DOCUMENT RESUME

ED 338 635 TM 014 438

AUTHOR	Rogosa, David; And Others
TITLE	Use of Achievement Measures in Judicial Decisions: Assessment of Student Academic Progress.
INSTITUTION	Stanford Univ., Calif. Inst. for Research on Educational Finance and Governance.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Aug 89
CONTRACT	NIE-G-83-0003
NOTE	267p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC11 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Gains; Achievement Tests; *Court Litigation; Decision Making; Educational Policy; Grouping (Instructional Purposes); Job Performance; *Longitudinal Studies; Racial Discrimination; School Desegregation; *Statistical Analysis; *Student Evaluation; Teacher Evaluation; Test Interpretation; *Test Use

ABSTRACT

The use of analyses of achievement test data, particularly longitudinal data, in a wide range of court cases bearing upon educational policy and practice is examined, with an emphasis on assessments of the educational progress of students. The main purpose of the project is to show that assessments of educational progress (i.e., student learning) are integral to many court cases and related educational policy issues. Project participants were educational researchers and specialists in quantitative methods. The major research questions addressed the types of questions about student progress that arise in educational court cases, the types of data and statistical summaries that are presented in court cases, courts' responses to the empirical evidence on student progress, and presentations that prove most successful. The representation of academic progress by student growth in achievement, the definition and representation of expected growth, and data and methods used in practice to assess student progress are discussed. Specific educational court cases involving the measurement of student progress are presented covering the evaluation of teacher performance, racial discrimination, mandated desegregation, and student classification. Reanalyses of the evidence presented in two cases are also described. Finally, suggestions are made for improved data collection, data management, and statistical analysis. A 61-item list of references, an annotated bibliography of additional sources, and two computer programs for analyzing student progress are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

e19149

USE OF ACHIEVEMENT MEASURES IN JUDICIAL DECISIONS:

ASSESSMENT OF STUDENT ACADEMIC PROGRESS

ED338535

David Rogosa
Evonne Schaeffer
Gary Williamson

Stanford University, School of Education

The research reported here was conducted with funds from the Institute for Research on Educational Finance and Governance (Grant NIE-G-83-0003). The analysis and conclusions do not necessarily represent the views or policies of either institute.

The authors wish to acknowledge the valuable assistance and advice of Professor Robert C. Calfee and his students in the conceptualization and conduct of this project.

TM 017438

BEST COPY AVAILABLE

ERIC

Use of Student Achievement Measures in Judicial Decisions: Assessment of Student Academic Progress

David Rogosa, Stanford University
Principal Investigator

EXECUTIVE SUMMARY

This research examines the use of analyses of achievement test data in a wide range of court cases bearing upon educational policy and practice. The particular focus of the project was on assessments of the educational progress of students. The participants in this project are educational researchers and specialists in quantitative methods. The aim was to study the use of empirical evidence in the court cases and to apply recent advances in the conceptualization and analysis of longitudinal data. A main purpose of the project is to show that assessments of educational progress (i.e., student learning) are integral to many court cases and related educational policy issues. Subsequent to identifying questions about student learning in a variety of court cases, the project sought to critique current practices and to offer prescriptions for improvement.

The major questions addressed in the project are:

* What kinds of questions about student progress in achievement arise in the educational court cases?

* What types of data and statistical summaries are presented in the court cases?

* What are the court's responses to the empirical evidence on student progress and what kinds of presentations would be most effective?

Although many important questions about student progress arise in the court cases, in none of the cases reviewed was a fully sound or satisfactory analysis of progress located. In many of the cases the student achievement data were seriously deficient, and in others the statistical analyses were weak or misleading (reflecting also the current state of practice in educational and behavioral science research). Recommendations for improved data collection designs are given, and improved statistical methods for the assessment of student progress are described which are also implemented in a set of computer programs.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

USE OF ACHIEVEMENT MEASURES IN JUDICIAL DECISIONS: ASSESSMENT OF STUDENT ACADEMIC PROGRESS

SECTION 0

PURPOSE AND OVERVIEW

This research project examines the use of analyses of achievement test data in a wide range of court cases bearing upon educational policy and practice. The project seeks to apply recent advances in the conceptualization and analysis of longitudinal data; thus, the focus is on the assessment of the educational progress of students. The work concentrated on uses of test scores for assessing student progress in court cases and on the presentation of improved and more appropriate methods and designs. A main purpose of the report is to show that assessments of educational progress (learning) are integral to many key judicial and related policy issues. Subsequent to identifying questions about student learning in a variety of court cases, the project sought to critique current practices and to offer prescriptions for improvement.

The major questions addressed in the project are:

- What kinds of questions about student progress in achievement arise in the educational court cases?
- What types of data and statistical summaries are presented in the court cases?

o What are the court's responses to the empirical evidence on student progress and what kinds of presentations would be most effective?

The participants in this project are educational researchers and specialists in quantitative methods. Thus, the emphasis of the work was on the kinds of empirical evidence that has or could be presented to the court, rather than on technical analysis of legal issues. This project is located in an interesting niche where interests and concerns of research statisticians, educational researchers, legal experts and education policy makers intersect.

The report is divided into four main sections. In the first section we (i) present the basic ideas of empirical assessment of student progress, (ii) outline some of the issues involved in definitions of "expected growth" or "normal educational progress", and (iii) conclude with descriptions of different types of test data and of the different types of statistical analyses used for the assessment of student progress. The second section reviews the different types of court cases examined in the first phase of this research. We identified four types (sometimes overlapping) of court cases in which assessments of student progress figure prominently: evaluation of teacher performance, racial discrimination cases, evaluation of school desegregation programs and student classification cases. The third section of the report focuses on methods for statistical analyses of student progress through the critique and reanalysis of two court cases which both raised basic issues about student progress and also whose data was obtainable for reanalysis. The fourth section attempts to explain what we feel are the valuable lessons learned from this research project. In particular, we offer recommendations for kinds of data and statistical analyses that would serve to provide useful evidence on the questions about student academic progress raised in many of these court cases. And second we offer a description of three kinds of important, but not fully understood, technical problems that would serve as a useful guide for future methodological research.

SECTION 1

ISSUES IN THE MEASUREMENT OF STUDENT PROGRESS

This section first presents the ideas and the data structure of individual histories of achievement and their use in statistical assessments of student progress. Next, the possible definitions of "expected growth" or "normal educational progress" are examined. Third, we consider the various forms and transformations of test scores that might be or have been used in assessments of student progress, and discuss the statistical methods used to summarize test scores. The aim is both to show the diversity of various methods that arise in the presentation of statistical arguments in relevant court cases and to establish a framework for thinking about useful methods for assessing student progress.

SECTION 1.1: Representing Academic Progress by Student Growth in Achievement.

School districts regularly assess student performance using group administered achievement tests. Such testing represents a large investment of money and time for the schools, for administrators, for teachers, and for students. Yet relatively little use is made of the accumulated test data. In particular, test results are presented in a way that describe only the current status of students; the data are presented as a static "snapshot" of achievement without any link to prior levels of performance. Even the management of test data reflect these limitations. Whether the test results be stored as hard copy or electronically, achievement data are organized as separate yearly files, which may be located on separate physical devices and even in separate geographical locations.

A key to the improved use of achievement test data is to use performance on repeated tests to describe student learning. A student's score at a single point in time cannot be used to measure learning: collecting together scores from previous testing is necessary for the analysis of student progress. A student's "cumulative folder" is organized in this manner, but these are rarely stored electronically nor uniformly maintained.

It is regrettable that achievement results tend to be presented within a static framework--the data for a given group of students are shown without any ties to their earlier performance and without any reference to the instructional program. Much important information is lost when test results are presented in a way that only informs about current behavior. A major obstacle to the improvement of the use of achievement test data is that school administrators and researchers or expert witnesses working with the school administrators do not have any practical models or methods to follow in bridging the gap between static snapshots of achievement data and assessment of student progress. Although questions about student progress or student learning are frequently seen in the court cases, these questions are rarely addressed adequately.

For improved use of achievement test data, the key is to use performance on the repeated tests to describe student learning processes. A basic belief underpinning the work of this project is that statistical analyses of individual achievement histories can be highly informative for description of student learning and for identification of strengths and weaknesses of instructional programs. By bringing together historical data from several years for individual students, useful indices for student progress can be calculated. This approach differs in two significant ways from methods currently used by practitioners and researchers. First, the student is the basic unit of observation and analysis. Rather than focusing on average scores which will obscure important individual differences, we begin with an analysis of each individual's record. Second, the emphasis is on multi-year trends rather than pre-post year to year changes, as information about progress or growth is extremely limited when only two points in time are available.

Though the notion of individual growth curves is not new in the study of learning and human development, educational researchers and policy analysts have given little attention to this important topic. A primary goal of this study is to advocate the creation and use of student histories of achievement, and to provide guidance for analyzing this information in the context of court cases.

Statistical Models for Individual Growth: Psychological learning theory and biological growth research provide a variety of complex models of individual growth such as polynomial growth curves, logistic growth curves, and simplex models (see Rogosa and Willett, 1985, for an exposition of these models). The simplest models, and the one we concentrate upon for practical application, is a straight line growth model which assumes a constant rate of learning for each individual. Thus, the rate of improvement provides a simple index for the learning of each individual. The straight line growth model is useful for heuristic reasons because of its simplicity and also serves as a useful approximation to actual growth processes. Moreover, the common use of grade equivalent scores for achievement test data which implicitly assumes a constant rate of change model with unit slope makes exposition of these kinds of models practicable to non-experts. Finally, the rather small number of observations typically available on each student (perhaps four) justify fitting models only as simple as the straight line growth curve model.

Figure 1 provides four examples of possible data for each individual to illustrate the data and the straight line growth model. For example, student A exhibits a rapid rate increase from a relatively low initial status. Student B starts out somewhat higher than student A and grows at a similar rate. Student C exhibits slow growth from a relatively low initial achievement level and finally student D shows a high level of achievement but no growth due to the ceiling effect of the test used.

- - - - - - - - - - -

Insert Figure 1 here

- - - - - - - - - - -

ERIC

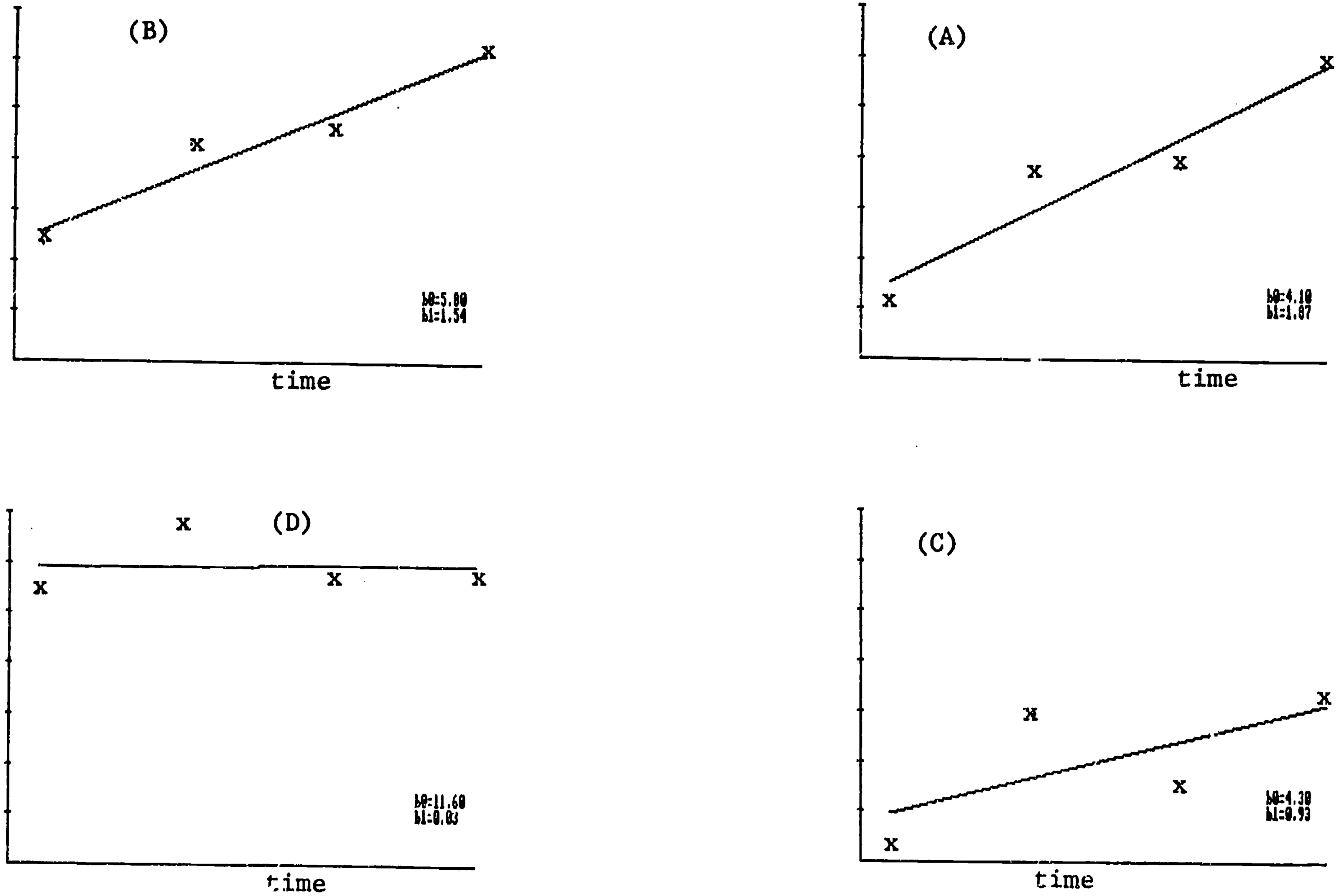


Figure 1. Contrastive growth trajectories for four individual students: (a) low entry, fast steady growth; (b) average entry, steady growth; (c)low entry, slow erratic progress; (d) high entry, ceiling effect.

ERIC

SECTION 1.2: Definition and Representation of Expected Growth

Comparison with a standard or criterion is a problem common to many uses of test data. Even with the static snapshot of achievement at a certain point in time, the interpretation of individual or group-averaged scores often rests upon the individual's standing relative to a national or local group (i.e., calculations of percentile rank scores) or the comparison of one group average with another (i.e., comparison of schools or district means with each other or with the state average.) Certainly, such concerns do not evaporate when indices of progress over time are at issue rather than snapshots at a single time. What is often desired after some measure of student progress has been calculated and or aggregated, is a comparative assessment of that index vis-a-vis some expectation. This common goal is a rationale for the preponderance of concern with normal educational progress or adequate progress that is seen in many of the court decisions. Determination of such standards or criteria, even with a sound statistical approach for assessing progress, is not unambiguous. The question, "what constitutes normal educational progress?" cannot be answered by technical considerations alone. Even if norms of measures of progress are available, it is unclear whether such criteria should be established unconditionally or conditionally. By "unconditionally" we mean a norm of progress that does not take into account student background, student initial status, or other possible influences on student progress for different types of students; an "unconditional" standard may not be the fairest or most persuasive in a judicial proceeding. Conditional criteria respond more to questions such as, "What is normal educational progress for a certain kind of student (e.g., ethnic minority)?" or, "is normal educational progress different for students who, say, entered high school with very different levels of achievement?" or, "is it adequate in evaluating a school desegregation program to determine that minority students have the same rate of progress as majority students even though a large preexisting gap between the level of achievement of minority and non-minority students was present before desegregation remedies were implemented?" As there is no natural metric, save that established artificially by grade equivalent scores, this important question has no simple or general answer.

But a crucial part of the use of analyses of student progess is a comparison with "expected" or "normal" growth. The best general advice is that in each specific application, explicit definitions of "adequate" progress should be made and deliberated among the parties involved. The two definitions of "normal educational progress" that are commonly used in court cases are based on either the grade equivalent score (Sections 1.3.1.4, 1.3.2.3) or the percentile rank score (Section 1.3.1.3). An improvement of 1.0 GE unit per school year is the most widely used criterion for normal progress. Alternatively, maintenance of percentile rank score over successive years is taken as "normal" growth (Section 1.3.2.3) in the sense that the student is moving along with the distribution of scores over time. In practice, these two criteria may frequently yield different conclusions for the same test scores.

Section 1.3: Data and methods used in practice to assess student progress.

The purpose of this section is to define and distinguish among the collection of data transformations of test scores and data analysis methods for the assessment of student progress that arise in the court cases. The first part of this section distinguishes among the various commonly used measures of scholastic achievement; that is, the various scores used to represent status of an examinee on a certain test or task. The most obvious of these is the raw score; more common are many complex transformations of the raw score. The second part of the section describes statistical analyses using one or more of the kinds of test scores for the assessment of scholastic progress. In the simplest case, there may be two measures, say, a year apart, available on each student and various statistical methods may then be applied to draw conclusions about the growth of groups of students and also to compare their growth with some sort of norm.

1.3.1. Different forms of test data.

This section describes raw scores and their common transformations: grade equivalent units, percentile ranks, normal curve equivalents (NCE), and scale scores.

1.3.1.1 Raw Scores.

The raw score is the simplest and most basic information on the examinee performance, being simply the total number of correct responses on the test or subtest. The raw score is the basis for all derived scores in that any score is based on a transformation of the raw score. The advantage of the raw score (or its equivalent, percent correct) is the simplicity of interpretation; a student taking the same form of the same test in two successive years may in the second year get seven more items right than she did in the first year. The seven item increase provides a simple idea of the student's improvement in scholastic improvement. For the static assessment of achievement level at a single point in time comparison of the raw score to the total number of items--which yields the percent correct--is obviously more interpretable in isolation than the raw score itself.

1.3.1.2 Normal Curve Equivalents (NCE). NCE's are a simple transformation of the raw score which serves in the attempt to give a readily interpretable static assessment of individual status for one point in time. If the raw score is denoted by X, the common form of the normal curve equivalent score is given by

$$\left(\frac{X - M}{S}\right)(21.06) + 50$$

where "21.06 was derived by dividing the distance from the mean to the 99th percentile...by the same distance measured in terms of normal curve standard deviation units." (Tallmadge, G.K. and Wood, C.W. (1976) User's Guide: ESEA Title 1 Evaluation and Reporting System. Mountain View:RMC Research Corporation.) That is, 21.06 = (99 - 50)/2.3267 = 49/2.3267. In the expression, M and S are the "estimated" national mean and standard deviation

(Tallmadge and Wood p.16) Thus, the NCE is just a simple linear transformation of the raw scores. In particular, for a single point in time, the correlation $r_{NCE,RAW}=1.0$.

In the definition of the NCE, the anchor points of percentiles 1 and 99 are arbitrary. For NCE's the multiplier (21.06) corresponds to anchor points of 1, 50, and 99. Define a as the area in one tail, so above a = .01. In general the anchor points are reflected in the multiplicative constant, $\{(1-2a)/2\}(100)/z_a$, where z_a is the upper (1-a) fractile in the N(0,1) distribution. So for example, if instead we want the transformed scale to be anchored at 5, 50, and 95 (a = .05), the scaling coefficient would be $100\{(1 - .10)/2\}/1.64 = 45/1.64 = 27.44$ rather than 21.06.

Although Raw scores and NCE's have a simple scale/translation relation at any single slice in time, use of the NCE metric for longitudinal data involves serious complexities. In particular, the norming constants (M and S) presumably will change at different points in time. Therefore the correspondence between raw score and NCE will not hold for assessments over time. One example of the use of NCE scores in a court case is described in section 2.4.2.1 (NAACP v. State of Georgia) where improvement in NCE scores in successive years for individual counties was used as evidence that student classification practices were not harmful.

ERIC

1.3.1.3 Percentile ranks.

Percentile ranks, one of the most common test score reporting metrics, simply indicate the percentage of raw scores in a norm group that fall below a given student's raw score. Thus, percentile ranks are a transformation of the raw score that is defined by the empirical cumulative distribution function of the norming sample. The form of this transformation is not linear but it is monitonically increasing with raw score. As an indication of relative standing to the norm group, the percentile rank has an obvious but limited interpretation for measuring achievement at a single point in time. For longitudinal assessment, the percentile rank score does not allow a simple interpretation of how much more a student knows at time 2 than at time 1, but percentile rank might be used to indicate whether the student is maintaining his or her standing relative to the norm group over time. As a percentile rank score cannot be used as an interval scale, more complex longitudinal statistical analyses of percentile ranks are not attractive.

1.3.1.4 Grade Equivalent Scores

Grade Equivalent scores (GE) are perhaps the most common and most abused scale for reporting results of standardized achievement tests. A grade equivalent represents the grade and month in school of students in the norm group whose test performance is theoretically equivalent to the test performance of a given student. A vast psychometric literature exists to discourage the overinterpretation of grade equivalent scores, and this literature serves a useful purpose. However, in urging that GE's not be over interpreted, sometimes the useful aspects of GE scores are not recognized.

The best way to think of GE scores is that like all other derived scores GE's are a transformation of the student's raw score on the test. However, the form of this transformation is usually more complex and less explicitly defined than for any of the other kinds of derived scores. A very large advantage of GE scores is their simplicity, both in reporting status in a single point in time and for the implicit characteristic of GE's in representing change over time. The simplicity of a unit increase per year of school has a compelling intuitive basis, if not always a sound psychometric basis.

Grade Equivalent scores are perhaps the most widely used metric for test data in court cases; for example, GE's are seen to be readily acceptable by the courts in the cases Tattnall County (Section 2.2.3) and NAACP v. the State of Georgia (Section 2.4.2.1) and Scheelhaase v. Woodbury (Section 2.1.1). If GE scores are to be deemed inadequate for technical reasons, then it may be necessary to have replacements with equally simple interpretations. Clearly additional complexity no matter its technical soundness is not necessarily efficacious in complex judicial proceedings.

1.3.1.5 Grade Equivalent Norms for Selected Tests.

In order to investigate the relationship between Grade Equivalent scores (GE) and raw scores (RS), we constructed plots of the GE's versus raw scores from norms tables for selected subtests. In general, we used a reading (or verbal achievement) score and a mathematics achievement score from each test battery. Table 1 shows the test batteries used, the specific level and form, and the recommended grade range for each. The Iowa Tests of Basic Skills (ITBS) were published by Houghton Mifflin. All the other tests were published by CTB/McGraw-Hill. These particular test batteries were chosen because of their wide use and the availability of norms tables.

\- - - - - - - - - - -

Insert Table 1 here

\- - - - - - - - - - -

Table 2 shows the maximum and minimum GE's and the raw score or range of raw scores that corresponds to those GE's for each test used. The names of the test batteries are abbreviated: CAT for California Achievement Tests, CTBS for Comprehensive Tests of Basic Skills and ITBS for Iowa Tests of Basic Skills. Each test battery abbreviation is followed by a slash and the designation for its level and form. Thus CAT/19D stands for California Achievement Tests, level 19, Form D. It can be seen that there is considerable variability in the correspondence between GE's and raw scores at the extremes of the scales. Some tests have one-to-one correspondences between GE and raw scores while others show a broad range of raw scores corresponding to the extreme

TABLE 1

TESTS ANALYZED IN FIGURES 1-16

TEST BATTERY	LEVEL	FORM	RECOMMENDED GRADE
CAT	19	C	9.6-12.9
CAT	19	D	9.6-12.9
CAT	5	A	9-12
CAT	3	A	4-6
CAT	2	A	2-4
ITBS	11	6	5
ITBS	14	6	9
CTBS	1	S	2.5-4.9

ERIC

GE's. Note particularly that the maximum GE on several tests can be achieved by answering less than 80% of the items correctly.

- - - - - - - - - - -

Insert Table 2 here

- - - - - - - - - - -

Figures 1-16 show the plots of grade equivalent versus raw score. The functional form of the GE-RS correspondence varies depending on the test. The CAT tests designed for use in the high school grades (9-12) generally show some ceiling and floor effects, as noted in Table 2 and Figures 1-6. In the intermediate range the scales show approximately a straight-line trend for the reading tests and a slight s-shape on the math tests. However even on the math tests, a straight-line might fit reasonably well for the middle range of scores since the curvature is slight.

- - - - - - - - - - -

Insert Figures 1-16 here

- - - - - - - - - - -

The CAT tests designed for the elementary grades (CAT/2A and CAT/3A) show a different pattern. There is some floor effect on each test (Figures 7-10). However, these tests show considerable curvature in the relationship between GE and raw scores with the curve taking a steep upward turn at roughly 3/4 of the maximum raw score. No ceiling effect is exhibited.

The GE and raw scores scales on the ITBS subtests (Figure 11-14) show roughly a linear relation throughout the entire range

TABLE 2

GE AND RAW SCORE RANGES FOR TESTS IN TABLE 1

TEST	MIN GE	RAW SCORE	MAX GE	RAW SCORE
CAT/19C Math	2.0	0-7	12.5	54-85
CAT/19D Math	2.0	0-7	12.5	53-85
CAT/5A Math	0.6	0-3	13.6	77-98
CAT/3A Math	0.6	0-10	11.0	108
CAT/2A Math	0.6	0-28	6.3	117
ITBS/11,6 Math	1.7	0	9.0	42
ITBS/14,6 Math	3.0	0	12.9	48
CTBS/15 Math	1.0	0-12	8.5	98
CAT/19C Reading	2.0	0-7	12.9	49-70
CAT/19D Reading	2.0	0-7	12.9	46-70
CAT/5A Reading	0.6	0-6	13.6	67-85
CAT/3A Reading	0.6	0-10	12.9	82
CAT/2A Reading	0.6	0-23	8.4	85
ITBS/11,6 Verbal	1.0	0	9.2	43
ITBS/14,6 Verbal	2.3	0	12.9	47
CTBS/15 Reading	1.0	0-9	9.9	83-85

Figure 1

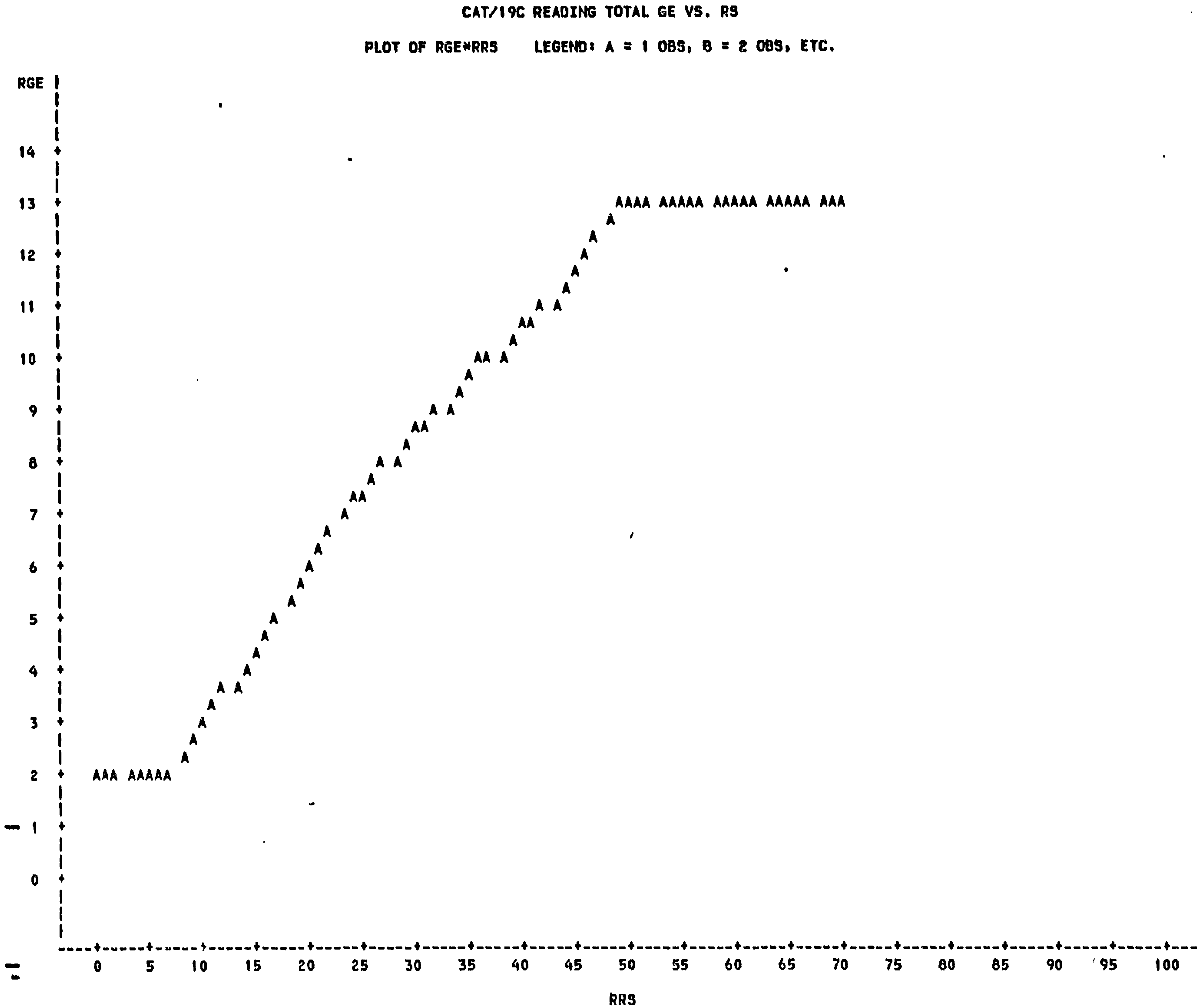

CAT/19C READING TOTAL GE VS. RS
PLOT OF RGE*RRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.
RGE
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
RRS

Figure 2

CAT/19C MATH TOTAL GE VS. RS

PLOT OF MGE*MRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

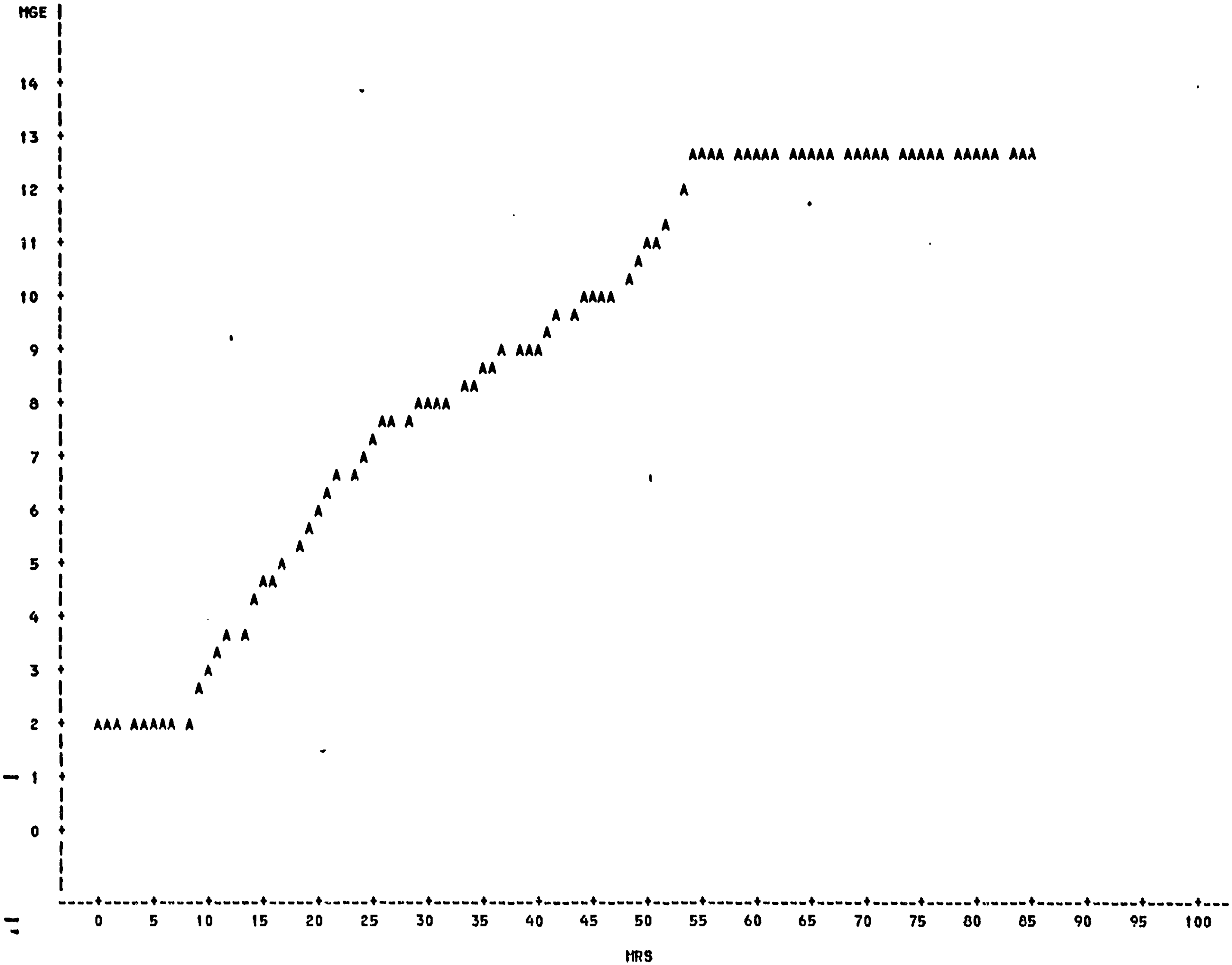


Figure 3

CAT/19D READING TOTAL GE VS. RS

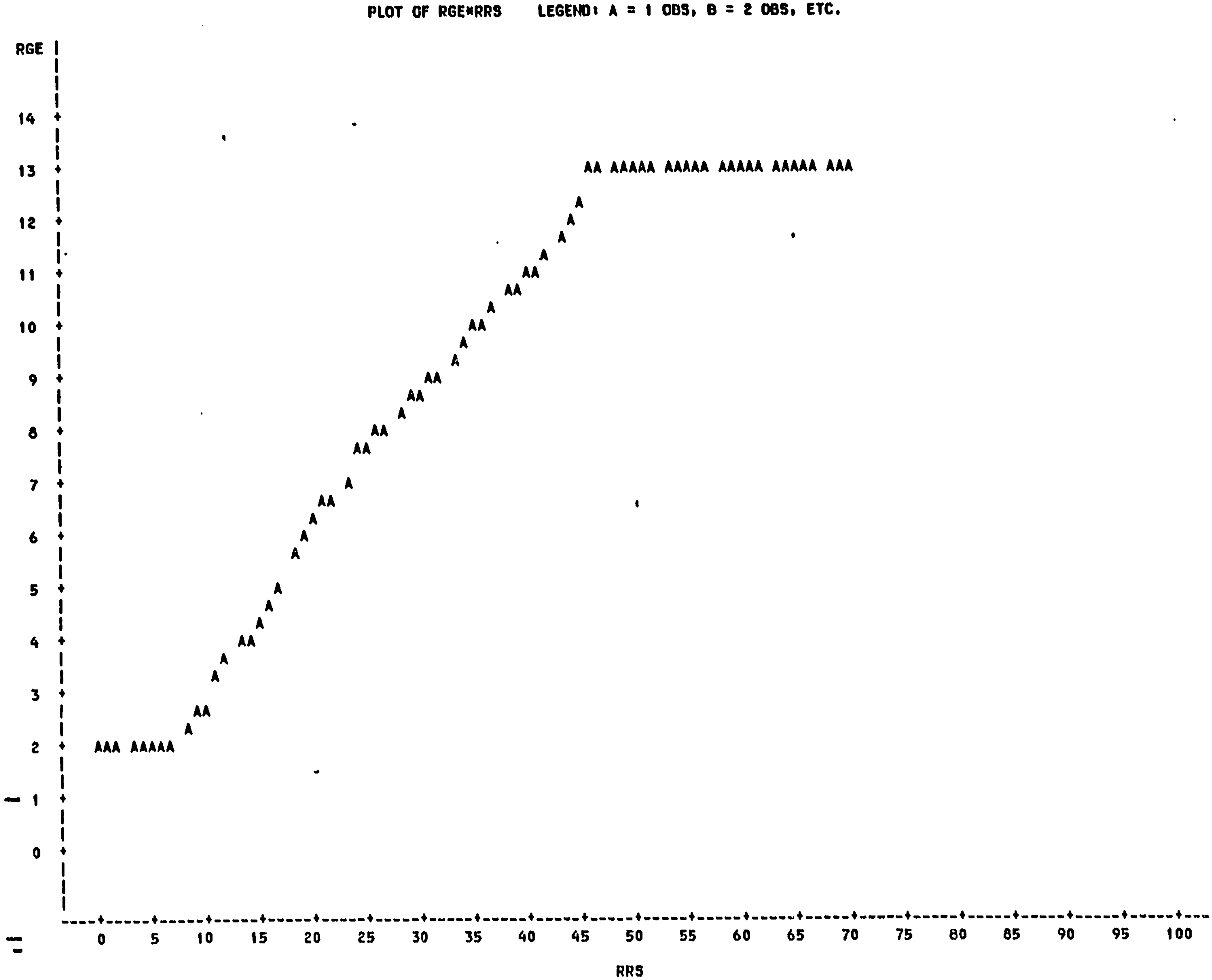


Figure 4

CAT/19D MATH TOTAL GE VS. RS

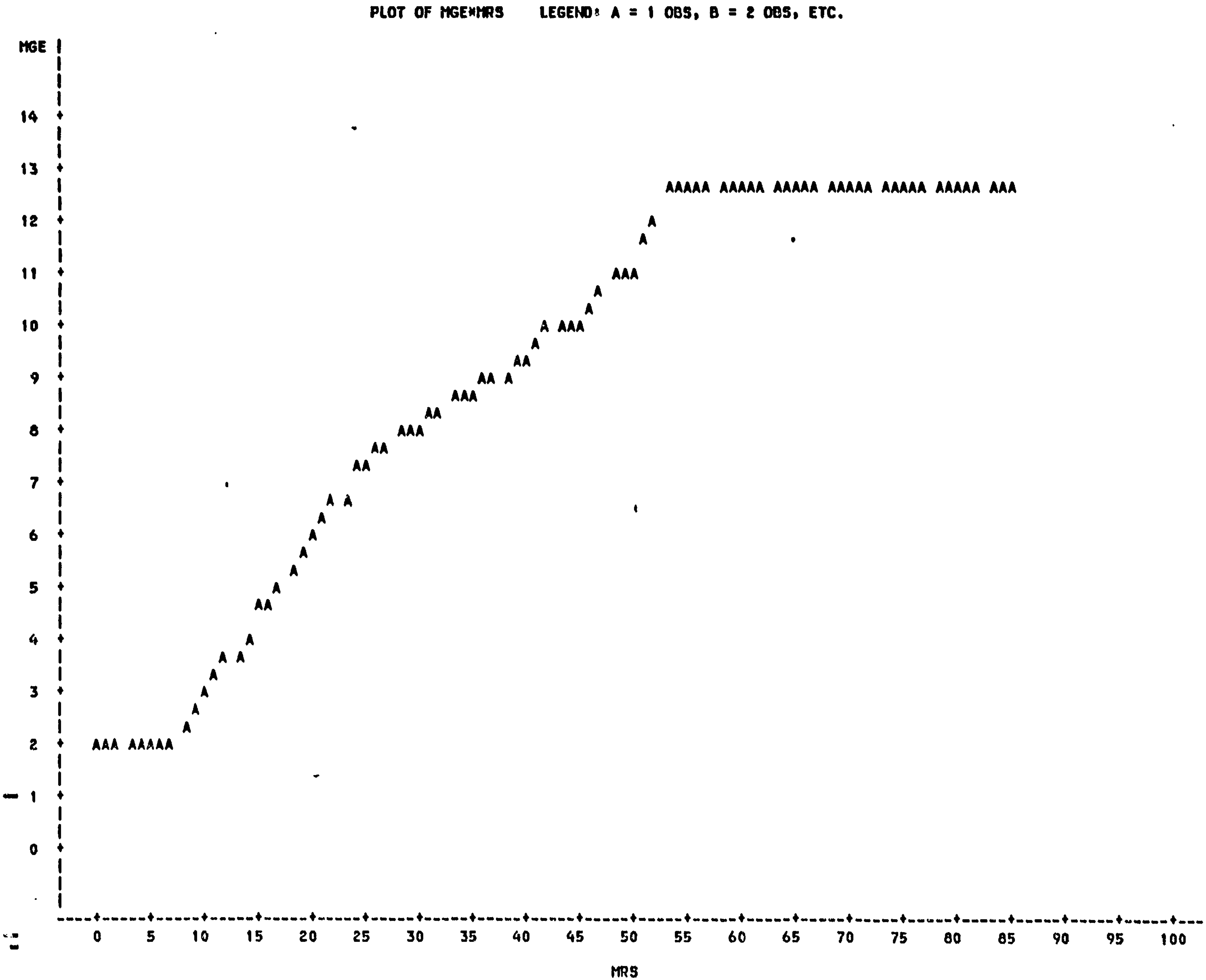


Figure 5

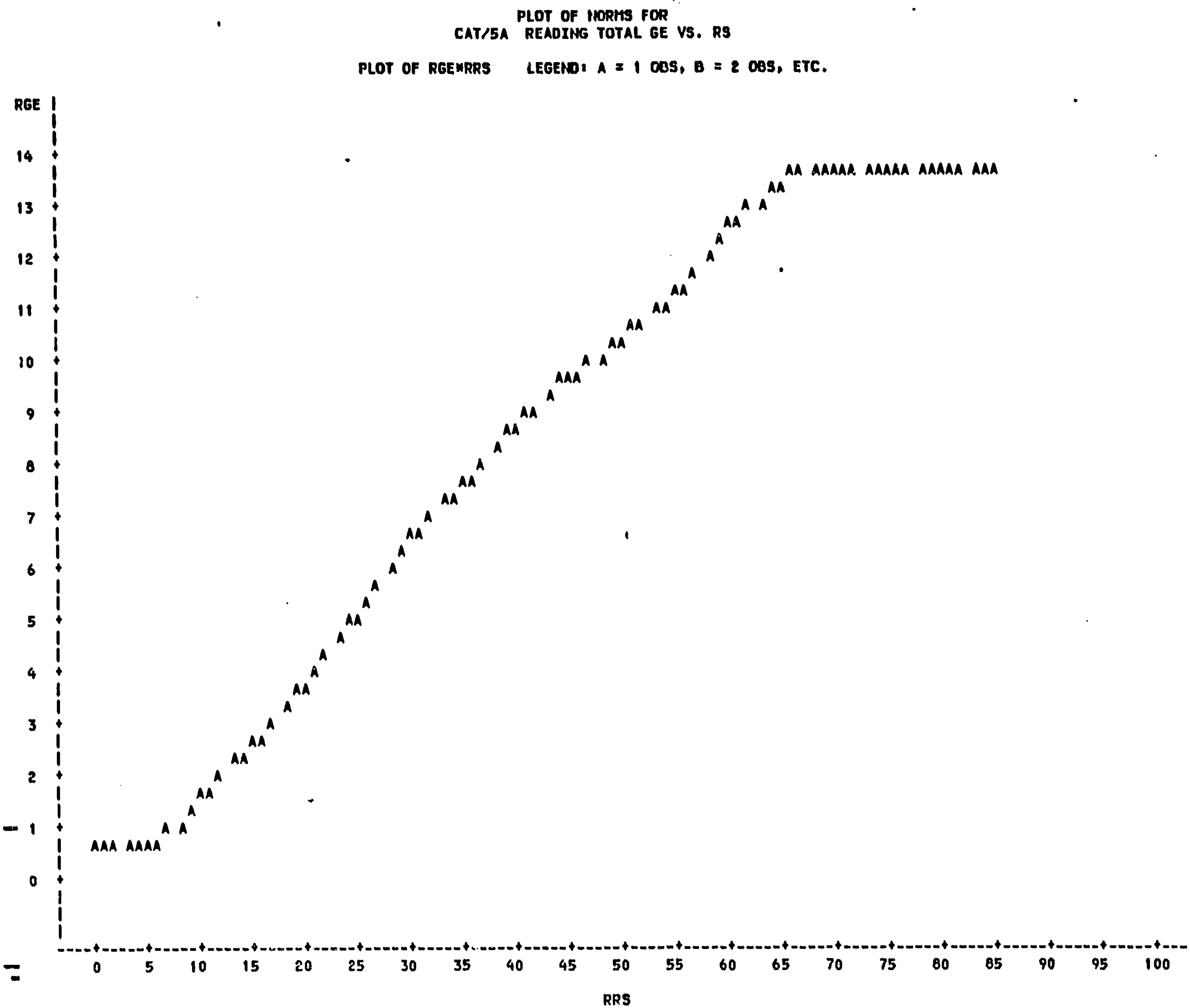

PLOT OF NORMS FOR
CAT/5A READING TOTAL GE VS. RS
PLOT OF RGE*RRS LEGEND: A = 1 OBS, B = 2 OBS, ETC.
RGE
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
RRS

Figure 6

PLOT OF NORMS FOR
CAT/5A MATHEMATICS TOTAL GE VS. RS

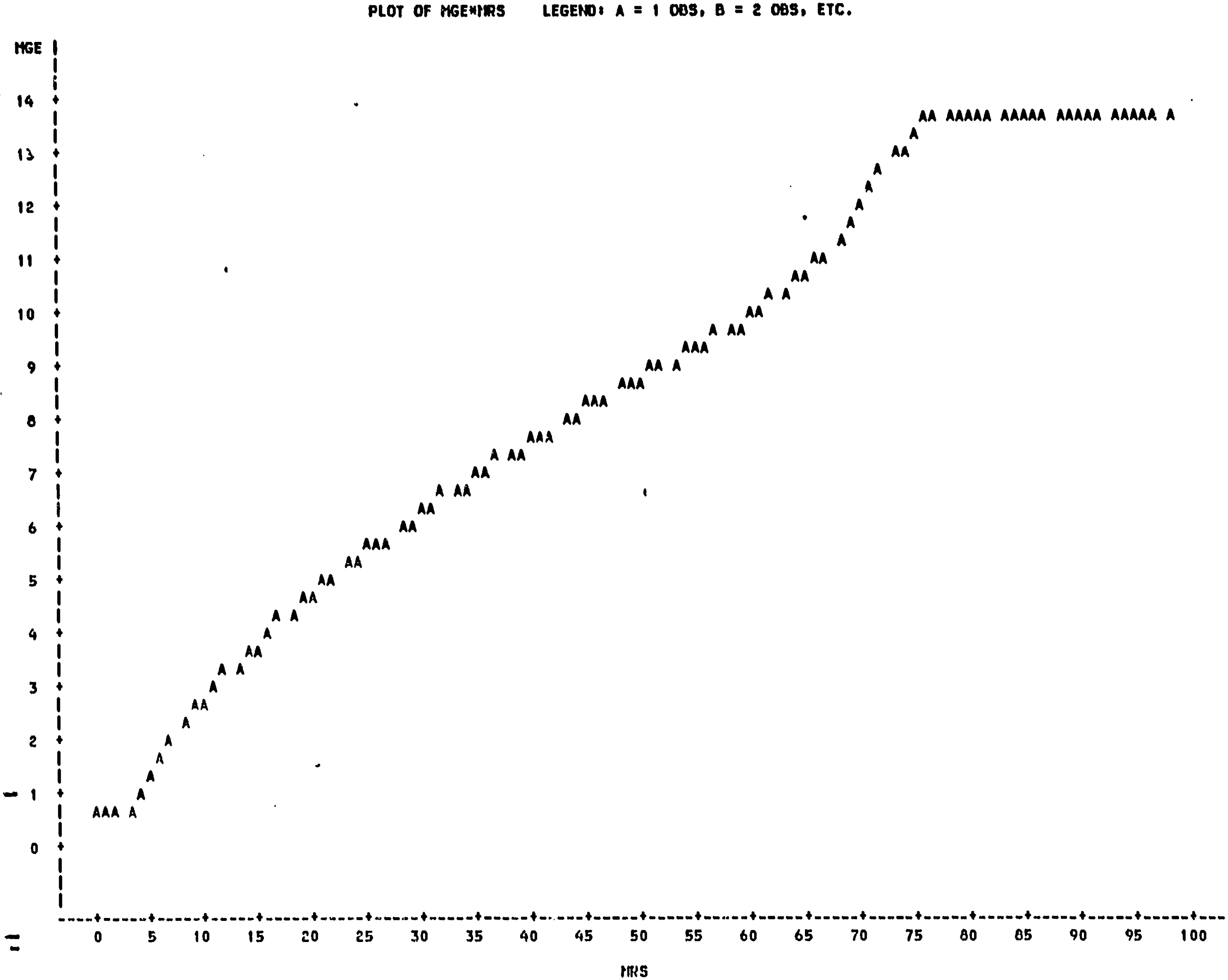
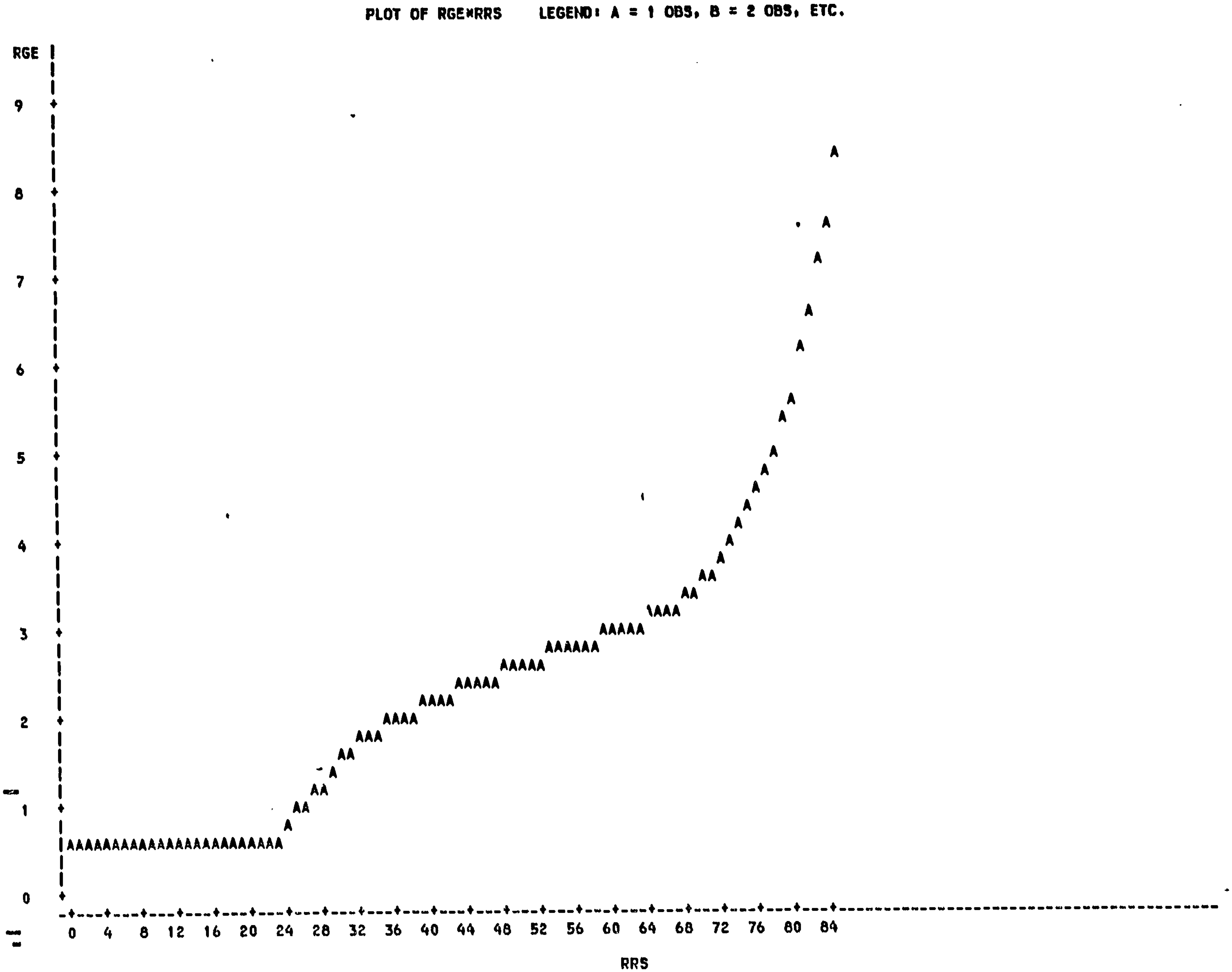


Figure 7

PLOT OF NORMS FOR
CAT/2A READING GE VS. RS.



PLOT OF NORMS FOR
CAT/2A MATHEMATICS GE VS. RS

PLOT OF MGE*MRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

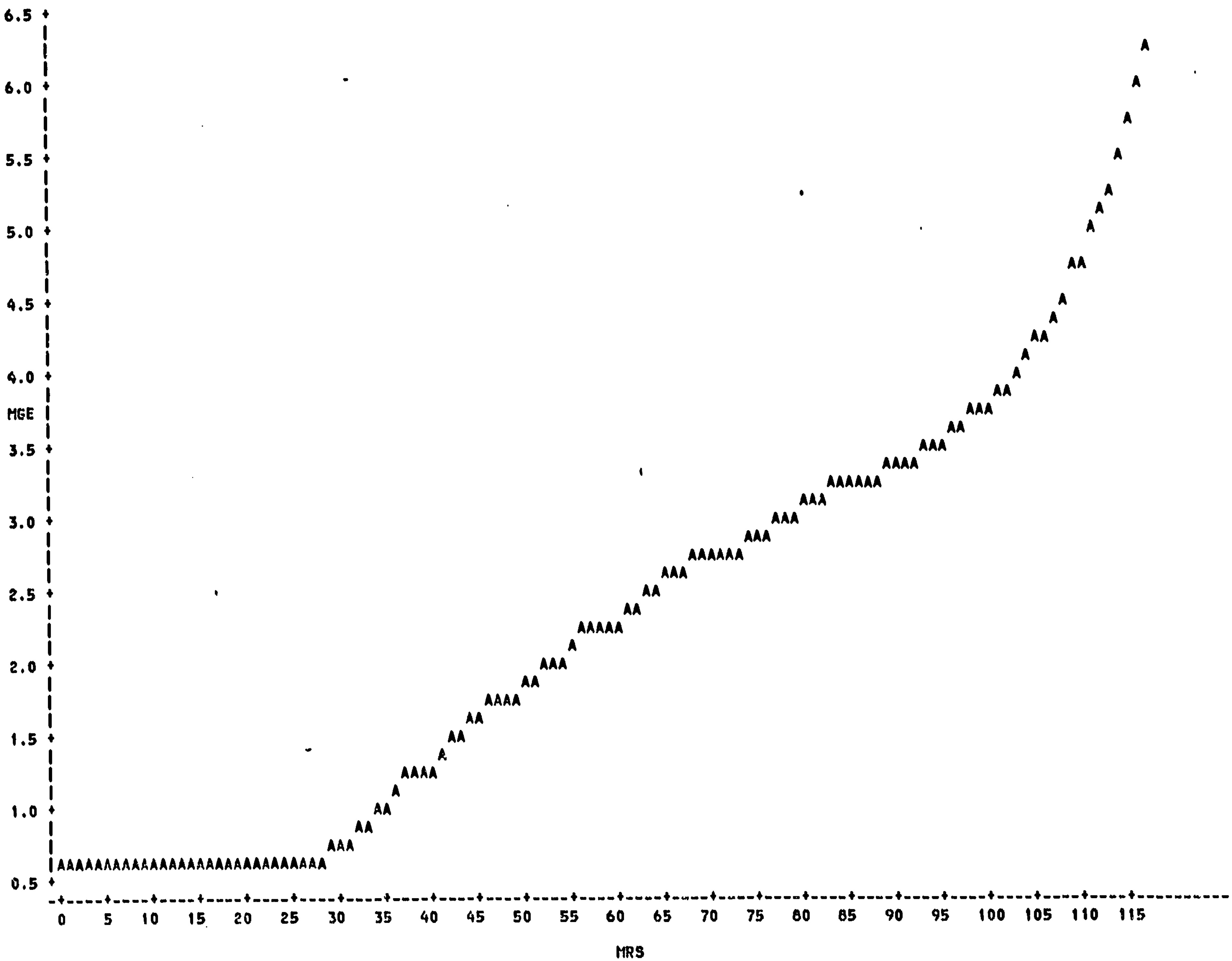


Figure 9

PLOT OF NORMS FOR
CAT/3A READING GS VS. RS

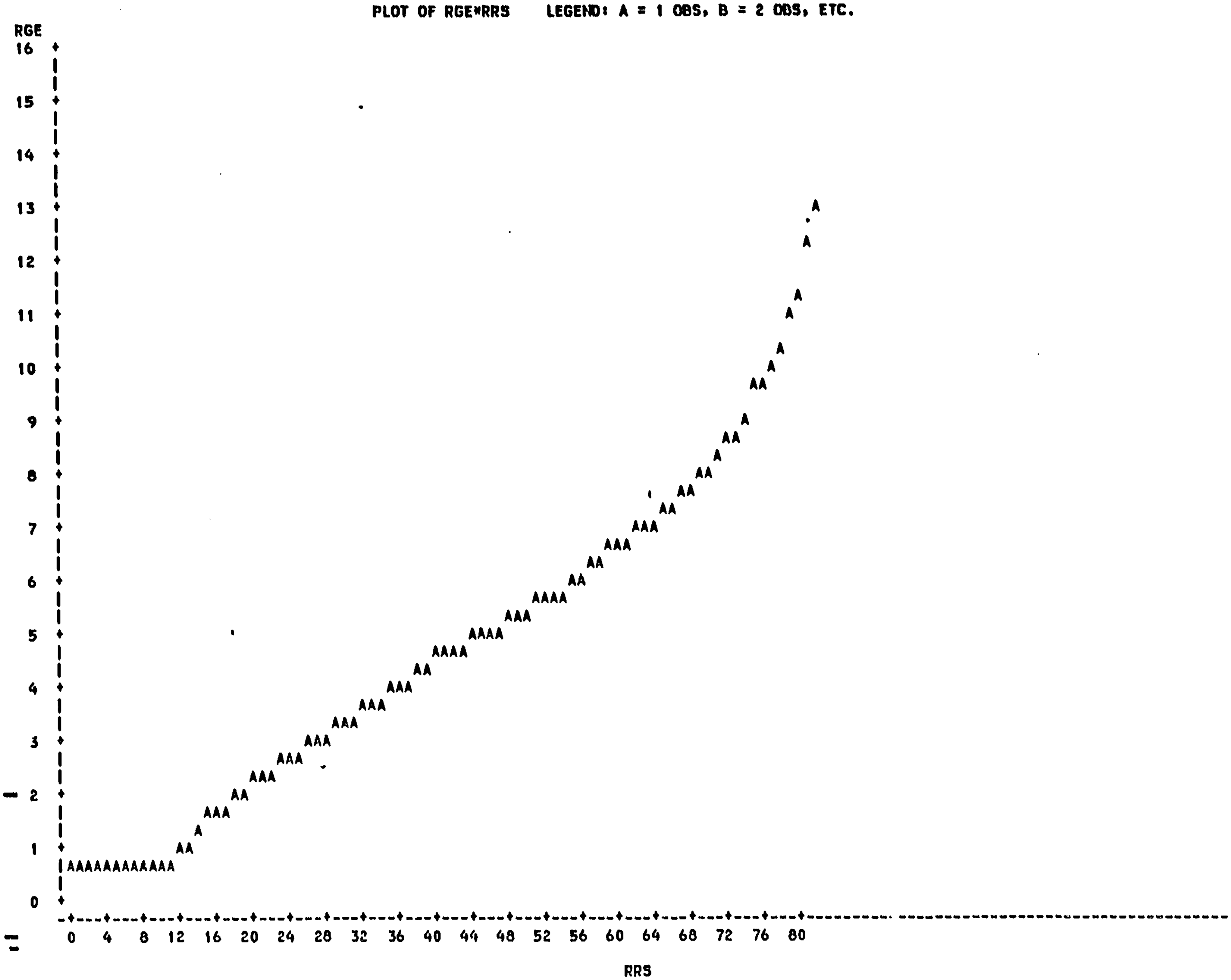


Figure 10

PLOT OF NORMS FOR
CAT/3A MATHEMATICS GE VS. RS

PLOT OF MGE*MRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

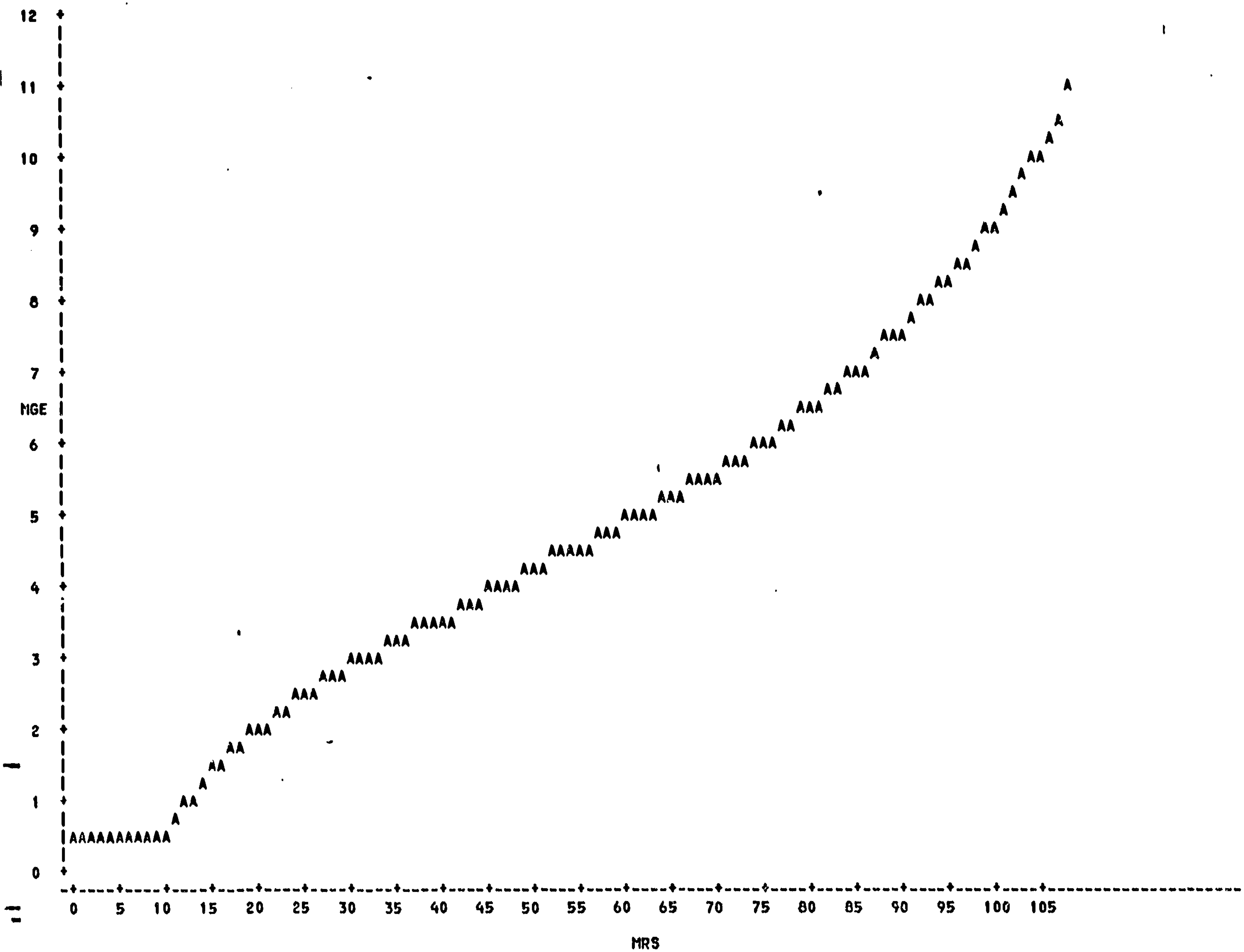


Figure 11

PLOT OF NORMS FOR
ITBS LEVEL 11 FORM 6 VERBAL GE VS. RS

PLOT OF VGE*VRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

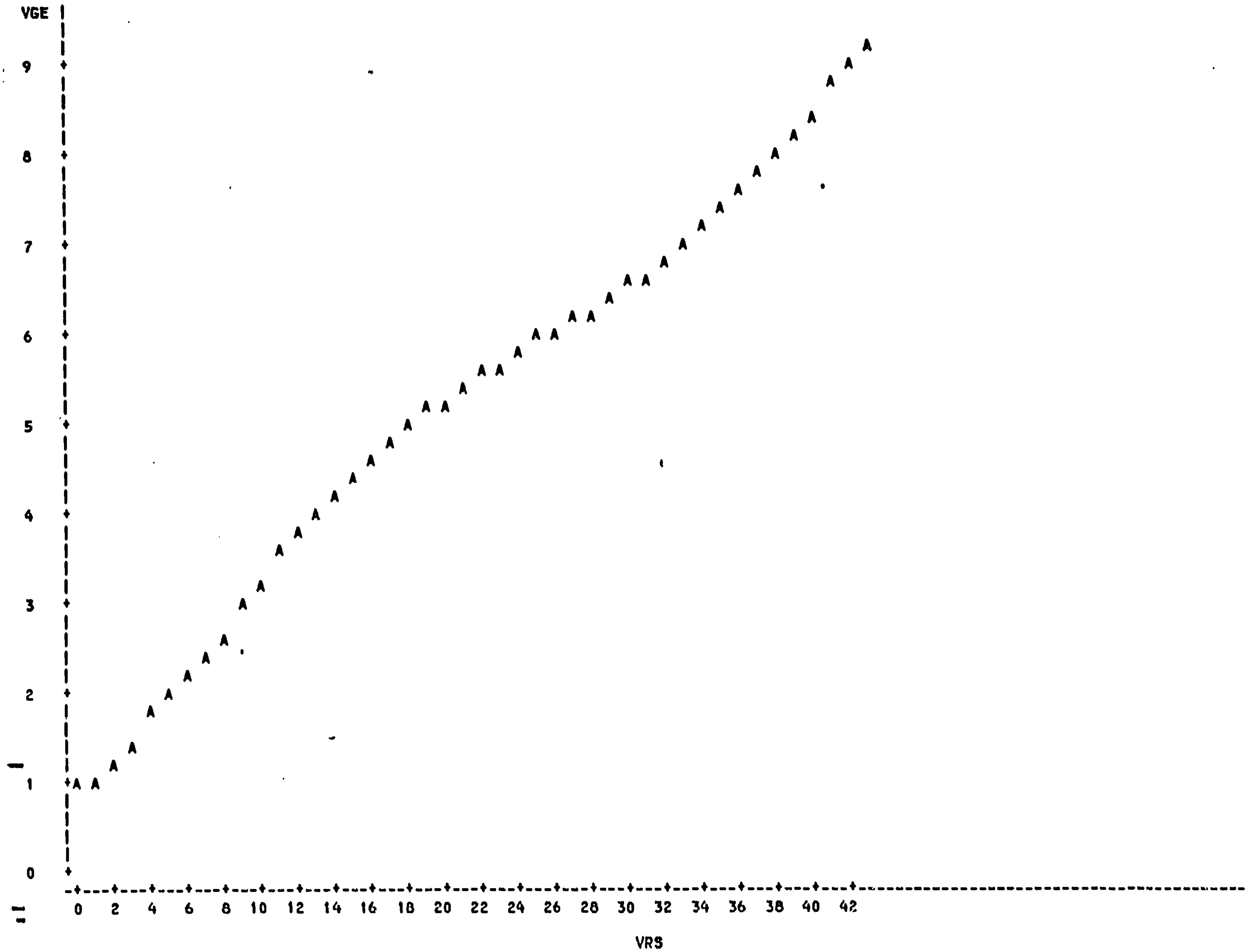


Figure 12

PLOT OF NORMS FOR
ITBS LEVEL 11 FORM 6 MATH GE VS. RS

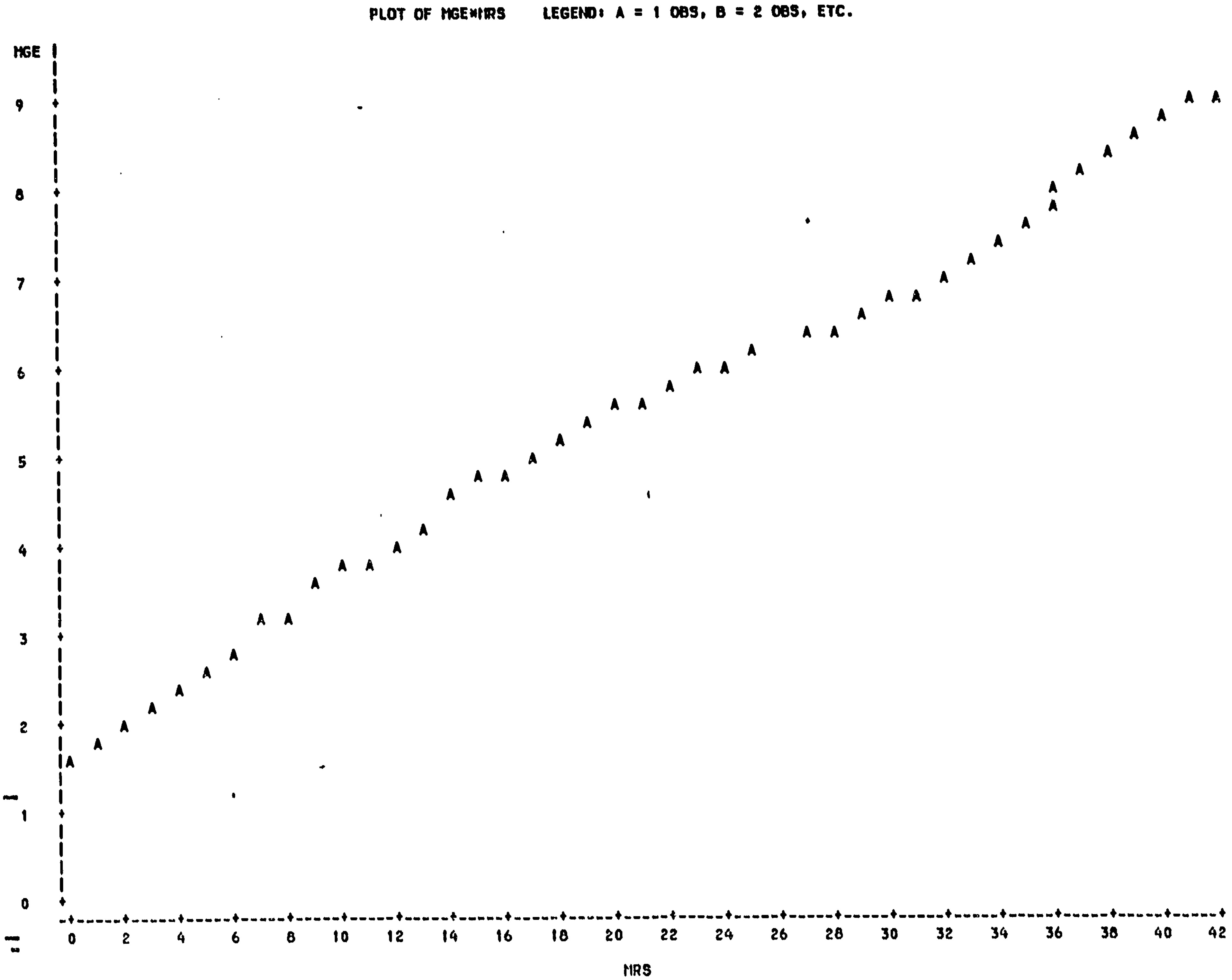


Figure 13

PLOT OF NORMS FOR
ITBS LEVEL 14 FORM 6 VERBAL GE VS. RS

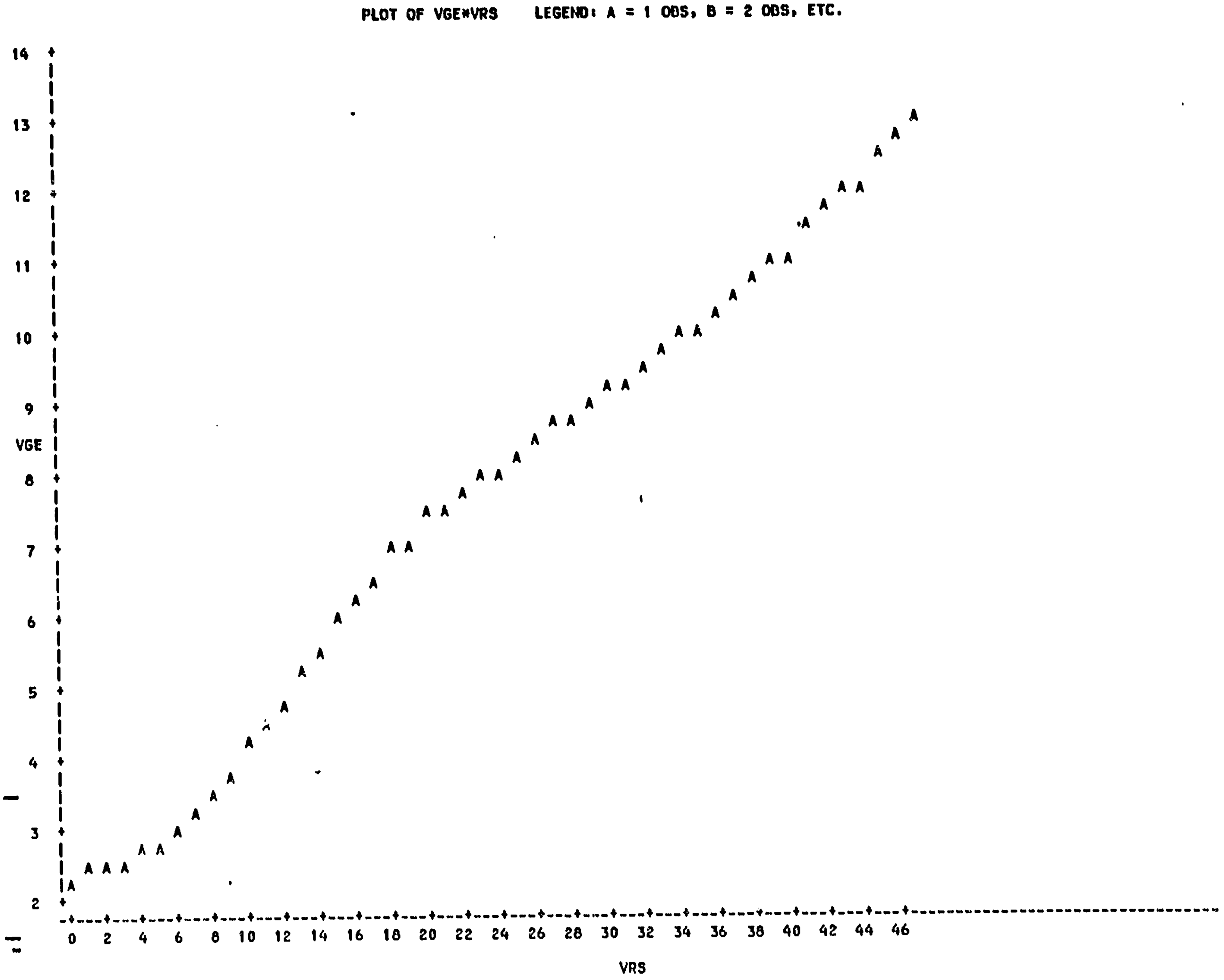


Figure 14

PLOT OF NORMS FOR
ITBS LEVEL 14 FORM 6 MATHEMATICS GE VS. RS

PLOT OF MGE*MRS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

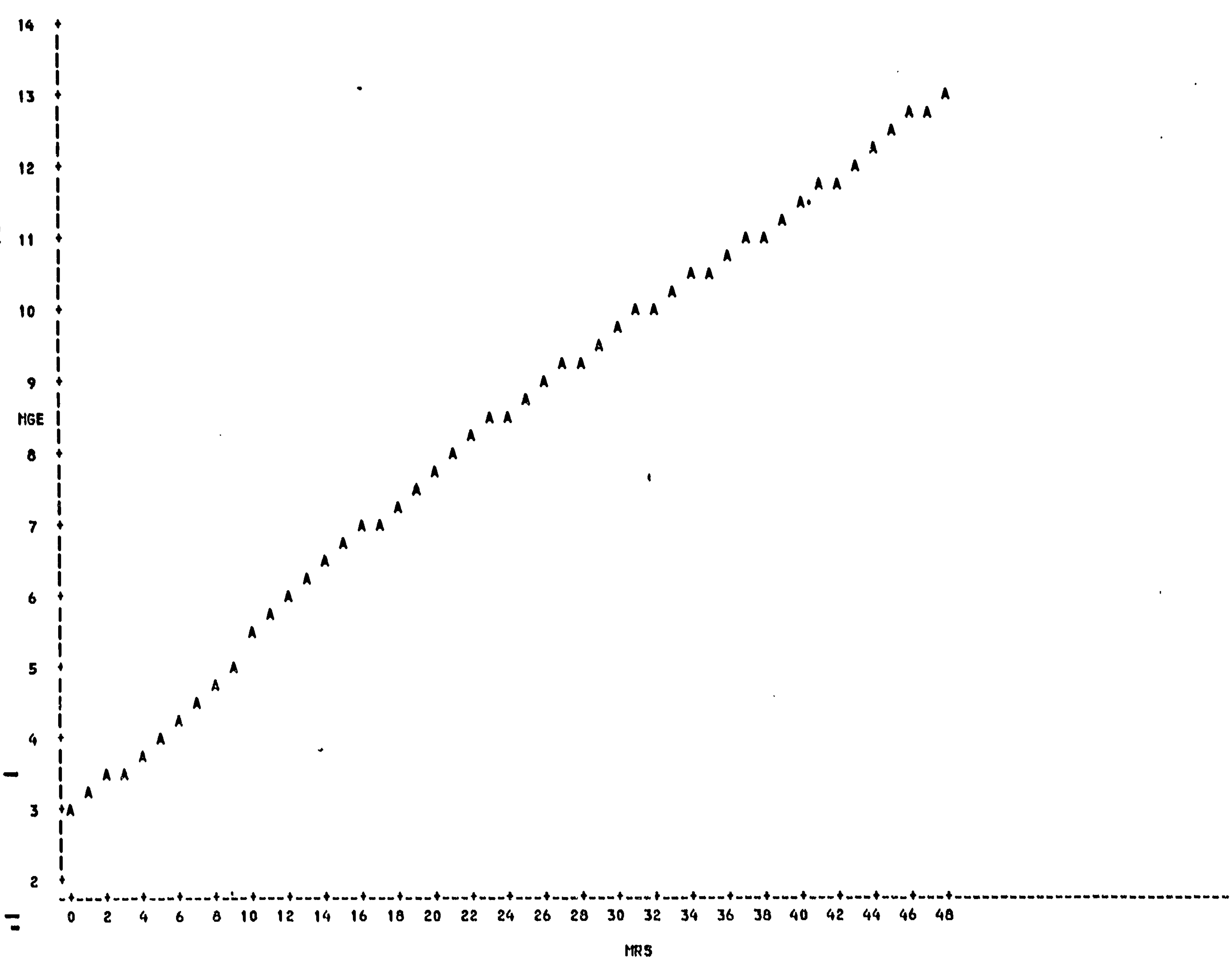


Figure 15

PLOT OF NORMS FOR
CTBS/1S MATHEMATICS GE VS. RS

PLOT OF MGE*MRS LEGEND: A = 1 OBS, B = 2 OBS, ETC.

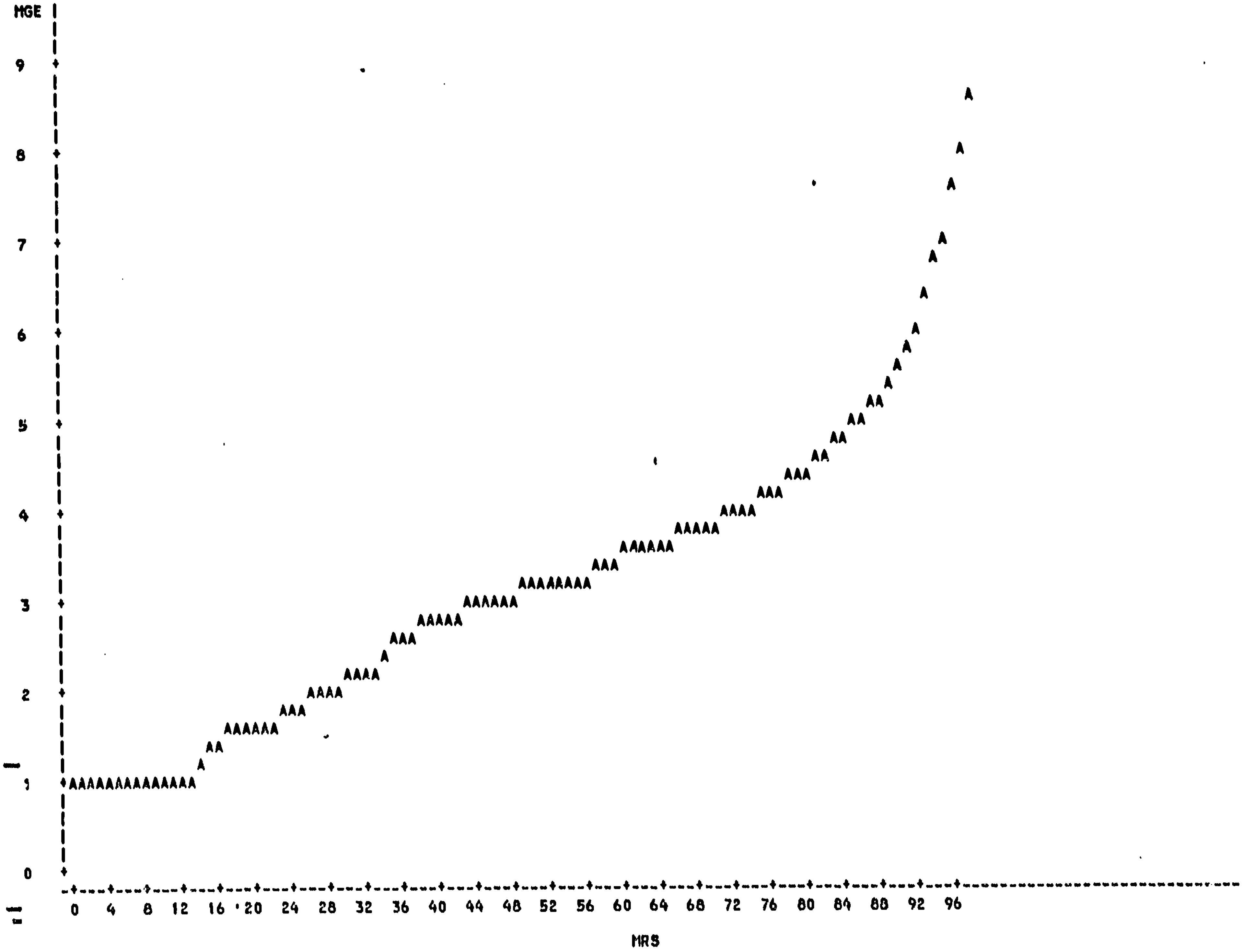
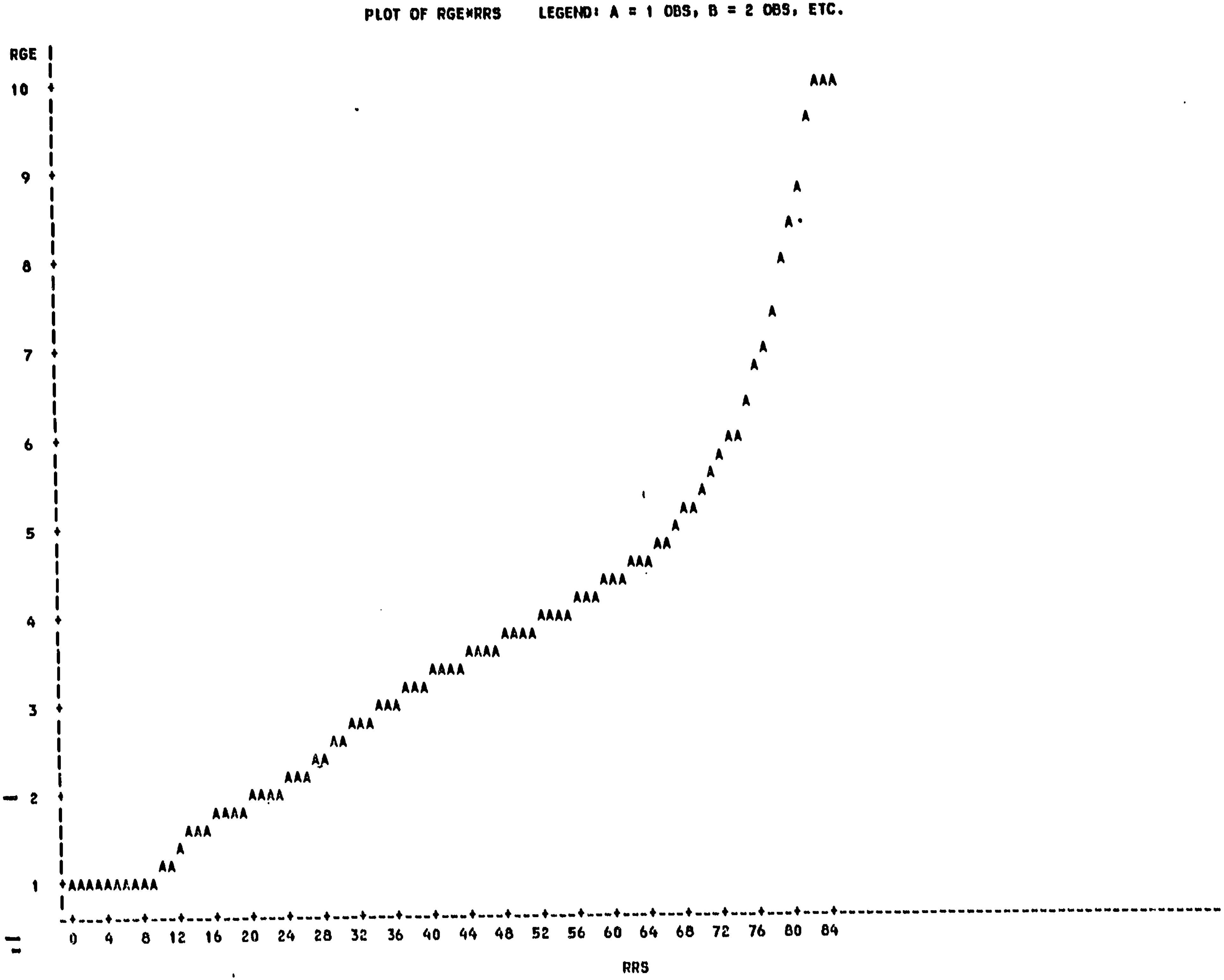


Figure 16

PLOT OF NORMS FOR
CTBS/1S READING GE VS. RS



of the scales with only a minor floor effect on the Verbal tests. The raw scores and GE's exhibit a 1-1 correspondence for the most part, with two raw scores occasionally mapping into a given GE.

The GE-raw scores relation on the CTBS/1S (Figures 15 & 16) is similar to that for the CAT/2A and CAT/3A, the other elementary grade products from CTB/McGraw-Hill. There is some floor effect. The curve then climbs slowly, approximately linearly, until it takes a steep upward climb at about 3/4 the maximum raw score. From that point on the curve climbs rapidly to the maximum point of the GE and raw scores scales.

1.3.1.6 Scale Scores.

Scale scores are the most recent and most sophisticated transformation of raw scores widely reported for standardized tests. These scores hold considerable progress in improving some of the undesirable features of other transformations of raw scores. The aim of scale scores is to provide a single equal interval scale of scores across all grades for use with all levels of a standardized test, such as the California Achievement Test (CAT). A key characteristic of scale scores is that they aim to be independant of test level or form, grade or time of year of testing. Also, the scale scores are intended to form an interval scale, that the difference of any two successive scores on the scale is always the same in terms of scholastic achievement. Thus, test publishers can claim the following relevant properties of test scores (taken from pg. 54 of the Test Coordinators Handbook of the California Achievement Tests for Forms C and D):

1. Scale scores can be added, subtracted, and averaged across test levels. Such computations make it possible to directly compare individuals, classes, schools, or entire districts.

2. Scale scores facilitate longitudinal studies of achievement growth over any period of time for an individual or group. For example, scale scores can be used to record a students progress from the beginning of school through Grade 12.

So, for purposes of assessing student progress, scale scores hold the most promise and could well overcome objections used with raw scores or grade equivalents in assessment of student progress. However, among the many court cases that were reviewed, only in some of the most recent San Francisco desegregation analyses (Section 2.3.2) were scale scores used; this is likely to be because test developers have only recently begun reporting such scores routinely. For example, CTBS tests were reported in scale score units starting in 1980. We would anticipate more use of scale scores in the future, and this should facilitate more useful analyses of student progress.

1.3.2. Methods for Assessing Student Progress.

Methods for assessing learning or progress in student achievement have been a major methodological issue in much of the educational and social science literature over the past 50 years.

Rogosa, Brandt, and Zimowski (1982) and Rogosa and Willett (1985) critique this literature. Some of these methods and their associated controversies appear in many of the court cases that were studied in this project. Almost any of the types or transformations of test scores described in section 1.3.1 can be used with the possible methods. However, some of these methods have much better or attractive properties with certain kinds of test scores than with others. This crossing of method and type of test scores produces a panoply of possible forms in which student progress might be represented. The purpose of this section is to describe some of these methods along with their use in certain court cases.

1.3.2.1 Gain Scores

The simplest, but often most controversial, method for assessing progress is simply to compute a gain or average gain between an earlier and later time for a certain test score metric. A gain score in raw score units simply relates how many more items the student (or an average of students) were able to achieve at time two compared to time one. The grade equivalent score has a particularly simple interpretation and natural linkage to the calculation of student gains. Grade equivalents attempt to provide an interval scale metric for calculating student gains expressed in terms of years of schooling. In addition, grade equivalents provide an implicit definition of expected gain in terms of a GE unit per school year representing some sort of average progress. Another obvious and perhaps better use of simple gains would be in terms of the scale scores (see section 1.3.1.5), which may provide a psychometricly more acceptable method for gauging student progress.

Most important, is that gain scores are limited to assessments of progress over two points in time--a limitation which can be extremely consequential. Assessments of progress may often require much more data than just two points in time on each student, but the difference score cannot make good use of this limited amount of data. When more extensive data are available, the difference score will discard important information about the form of growth or about nonconstant rate of change.

Examples of the use of gain scores abound in the court cases. For example, in the case NAACP v. State of Georgia (Section 2.4.2.1) the defendant school districts expressed their evidence of student progress in terms of gains over a school year both in units of normal curve equivalents and in units of grade equivalents. In Scheelhaase v. Woodbury (sections 2.1.1 and 3.2) student progress was summarized by different scores of grade equivalents taken at yearly intervals such as between sixth and seventh grade.

Although the basic notion of looking at improvement from one grade to the next via a difference score is simple and compelling, it is easy to see that the variety of forms in which test scores could be reported could yield rather different results from the same basic data. Possible contradictions can be illustrated through use of the plots of raw score versus Grade Equivalent scores accompanying section 1.3.1.4. The fact that the transformation of raw scores into Grade Equivalents is non-linear could mean that a sizeable increase in raw score units might translate into a very small Grade Equivalent increase in, say, the higher percentage correct part of the scale. Yet, on the other hand, for the low scoring students a small increase in number correct on the test can translate into a large apparent gain in Grade Equivalent units. Even less attractive in the assessment of progress by difference scores is the use of normal curve equivalents; because these scores are independently referenced to a target population at each separate time point, the comparability of the raw score units from which they are derived at two different times is destroyed.

1.3.2.2. Regression Adjustments.

Various forms of multiple regression procedures are employed in the court cases with two broad purposes: (1) prediction of later achievement with special attention to the importance of demographic or racial variables in the prediction and (2) regression adjustments for initial differences on achievement or demographic characteristics with the purpose of comparing change across initially dissimilar groups.

Prediction equations. A typical example of the first type of multiple regression appears in NAACP v. Georgia (section 2.4.1.1). In the regression analysis the outcome variable was tenth grade achievement test score in either reading or math and the predictor variables included prior achievement level, racial identification of the student, status in free/reduced lunch program (an SES indicator), student retention history, student absences, and remedial courses taken in mathematics and language arts. Basically, out of a pool of possible predictors, the predictors which show the highest statistical significance are culled and interpreted. In this particular analysis, prior test score, student race, and the free/reduced lunch indicator were found to be "important" while the other variables were found to be "insignificant". As summarized in the district court judgement, "plaintiffs conclude from this particular analysis that the discriminatory treatment which produces classroom segregation in the lower grades also produces differences in BST [achievement] performance on the tenth grade level" (CV482-233 U.S. District Court, Georgia p.32).

The usual interpretation of this kind of multiple regression analysis focuses on the notion that if a variable such as racial identification of the student appears important in the prediction of later achievement with prior achievement also used as a predictor, then the data indicate that different racial groups have different rates of academic progress. The question, Do blacks and whites progress at different rates? is a typical question regarding a correlate of change--that is, Is racial identity a correlate of change in achievement? Rogosa and Willett (1985), in a technical critique and review of various methods for assessing correlates of change, find that these kinds of regression methods have very serious shortcomings and cannot be depended upon to identify correlates of change when such exists.

Adjustment for differences in initial status. A distinct, but conceptually similar, use of regression methods arises in adjustments of later achievement levels or to gain in achievement for differences in initial status in variables such as initial achievement or demographic characteristics. The most common form of such adjustments comes in the construction of residualized change scores. Such scores attempt to answer the question, How much would the students have changed if everyone had started out equal? or, equivalently, What would be differences in later status if everyone (i.e. different racial groups) had started out with equivalent initial status? A motivation for the use of such regression adjustments in both behavioral science research and in various court cases arises from concern that observed progress (e.g. the difference between later and prior achievement levels) is confounded or correlated with initial status. For example, in the Tattnall County case (sections 2.2.3, and 3.1) it was shown that "students who made the largest gains in their CAT scores during their high school years were the students who scored at the lowest levels on entry to high school."(540 F.Supp. at 764). It is often assumed that this kind of result is an artifact rather than an actual indication of differential student progress, and concern with such an artifact is used as justification for the construction of the residualized change score. Rogosa et al. (1982) demonstrate the highly unattractive properties of residualized change scores as a measure of progress, and Rogosa and Willett (1985) further demonstrate that correlations of residualized change scores with background variables such as racial identification or demographic characteristics serve very poorly to indicate variables that are associated with differential growth or progress.

1.3.2.3. Comparison with a standard for student progress.

Of most consequence in assessing student progress is to interpret or judge the amount of student progress that the data indicate by comparison with a standard for some sort of expected progress. A question often asked is not "are students learning?" but, "are students learning enough?" The grade equivalent method of one unit per year of schooling is the most obvious example of such a standard and this appears in Scheelhaase v. Woodbury (section 2.1.1), NAACP v. Georgia (section 2.4.1.1), and San Francisco NAACP v. S.F.U.S.D. (section 2.3.2). Also, in NAACP v. Georgia, a standard of expected gain of one Normal Curve Equivalent per year is also presented by the defendant school districts. "In 1978, when the expected gain in number systems, computation, and application was 1 NCE, only three students from among over 200 tested failed to meet this expected gain" (CV482-233 page 23). In Berry v. Benton Harbor comparison of achievement level with norms for the state of Michigan was used; such a comparison with a relevant group, when such exists, is another common form of assessing progress relative to a standard. Yet another implicit standard for progress is often attempted by use of test scores in the metric of percentile ranks. In particular, a notion of normal educational progress could be formulated by the standard of maintenance of percentile rank standing over successive years. Of course, this kind of standard raises the question of whether adequate student progress exists if a low scoring student at say, the 30 percentile rank manages to maintain that percentile rank over the course of school years. This is a controversial standard as many courts hope that educational programs, such as desegregation programs, will tend to decrease the gap between minority and majority students, not just allow the maintenance of such disparity.

Regression adjustments to growth rate. In the Tattnall County case (section 2.2.3) analyses were presented that combined growth curve methods and regression adjustments for differences in initial status in a determination of whether student progress was "greater or less than expected." The analysis begins with the estimation of a straight-line growth curve for each individual using achievement scores at the four high school grades. The estimation yields for each individual a slope (rate of progress) and an intercept (initial status at beginning of high school). Then, from the (rate, intercept) pairs for each individual, a scatterplot is formed of rate (ordinate) versus intercept (abscissa). Superimposed on this scatterplot is a straight-line fit of rate on intercept. Points (individuals) above the line are deemed to be making progress better than expected and those points below this regression line are deemed making progress worse than expected. Attention is also given to comparing this determination for individuals above and below the median on initial status; the regression line and median split on initial status divide the scatterplot into four quadrants. Rogosa et al. (1982, p. 743) show the equivalence between residual change procedures and this type of regression adjustment to the estimated growth rate.

Student progress in ability grouping. In court cases revolving around the adequacy or non-discriminatory effects of ability grouping or tracking a common question is, "Are students making adequate progress in each of the tracks?" One commonly used method is described by Findley and Bryan (1975). Findley and Bryan present data from a southern school district on the amount of change students in varying ability-grouped levels attained over two years. The twenty-four comparisons within each ability level consist of eight subtest scores for each of three groups. "A group was said to have made better than expected progress on a test if the percentile equivalent on the national norms of the mean at the higher grade was two or more points higher than the percentile equivalent of the mean at the lower grade had been for that grade; likewise, a group was judged to have made less than expected progress if the percentile equivalent of the mean on the national norms at the higher grade was two or more points lower than the percentile equivalent of the mean at the lower grade had been for that grade" (p.14). Table 3 reproduces their example which shows an apparent strong effect of tracking which is positive for high tracks and negative for low tracks.

- - - - - - - - - -

Insert Table 3 here

- - - - - - - - - -

The notion of evaluating the tracking system by measuring whether those in lower tacks maintained normal progress was also found in Mills and Bryan (1976). "To determine whether ability grouping in a school system is beneficial or detrimental to students, HEW compares the achievement gain or loss for students in the low group with the achievement gain or loss for students in other sections--middle, high, etc.--or the same grade at the same school. If, for example, the slow group at a particular school ranked at the thirtieth percentile when compared with the national norms at the third grade and ranked at the fifteenth percentile when tested at the fifth grade, while the middle group at the same school ranked at the fiftieth percentile in both third and fifth grades, then the grouping has not resulted in

TABLE 3

CHANGE IN PERCENTILE RANK FOR FOUR LEVELS OF ABILITY GROUPING

(adapted from Findley and Bryan, 1975, p. 14)

Level	Better than Expected	Expected Progress	Less Than Expected
I (high)	18	0	6
II	13	5	6
III	5	7	12
IV (low)	2	7	15

equal educational benefit for the two groups" (Mills and Bryan, p. 43).

1.3.2.4. Monitoring Student progress.

Another form of the assessment of student progress, denoted by the term "monitoring," uses much richer and much denser data collection stategies than is typically represented by yearly standardized achievement testing. One of the most prominent examples of monitoring of student progress arises in special education contexts reflecting the requirement for Individualized Educational Programs (IEP) mandated by Public Law 94-142. One quantitatively based system for monitoring student progress that is used especially in the Northwest is the Precision Teaching Model (described in White and Harring, 1980). The Precision Teaching Model provides a way to account for individual student progress. A fundamental assumption underlying the Precision Teaching Model is that learning is presumed to occur at an exponential rate. That means that achievement plotted against time will have a constant rate of increase if achievement is measured on a logarithmic scale rather than in the actual raw score or percent correct. The Precision Teaching Model provides an interesting departure from the constant rate of change models implicit in grade equivalent scores.

This accounting method begins by identifying a specific objective (e.g., 90 percent correct on a spelling test of fiveletter common words), and a certain date for this objective to be reached. Initial status (e.g., score on the same type of test) is plotted at the beginning date. The special graph paper is logarithmic on its ordinate, such that a straight-line on this paper represents a log function. The "minimum acceleration line" is drawn, connecting the objective with initial status (across time). The student must progress at least at the same rate as the minimum acceleration line to reach the goal by a given date.

The student's daily progress is plotted on the graph. If the observed points are lower than the expected rate for three consecutive days, a new intervention (i.e., different learning materials or methods) is implemented. If the student continues to fall below the expected line, the objective may be evaluated, restated, and even redefined.

No attempt is made to compare progress across students. Rarely is the goal for a student to join regular classes, although, for borderline cases, often a student is in regular classes in the morning, for example, and special reading classes in the afternoon. Mobility across special education classifications (EMR, LD, ED,...) or between special or regular classes is not common, and is rarely specified as an objective.

1.3.2.5. Improvement in classification status.

An important but qualitatively different kind of student progress arises in court cases involving student classification: either into ability groups or classification of students into special education classes. The student progress being assessed is not an improvement in achievement but whether or not changes in classification are non-discriminatory or appropriate. Previous issues regarding ability grouping ask the question, "Given the ability grouping that exists, are students making adequate progress?" Another aspect of that question is, "Are students reclassified into appropriate groups if they have a certain level of progress or educational improvement?" In the Dillon County case (section 2.4.1.3) the ability grouping of students was found to produce highly differential mobility patterns for black and white students. In cases such as Hoffman v. New York (Section 2.4.2.2) a key issue was whether classification into special education versus regular classes was sensitive to the progress of the student.

SECTION 2

EXAMPLES OF EDUCATIONAL COURT CASES
INVOLVING MEASUREMENT OF STUDENT PROGRESS

The court cases involving measurement of student progress identified in our research are organized into four major areas which we term evaluations of teacher performance, racial discrimination, evaluations of school desegregation programs, and student classification programs. Major purposes of this research project are to catalog and better understand the types of data and summaries that have been used, and to suggest improvements in both design and analysis. Relevant cases were identified by searching descriptive phrases which provide relevant key numbers in the West's Digest and Federal Digest systems. Most cases were located under the descriptive phrases of: educational progress, student progress, student test scores, and student achievement.

Evaluations of Teacher Performance: Court cases involving the evaluation of teacher performance are becoming increasingly visible due to various state and federal interests in teacher accountability and competence. We would expect court cases regarding teacher performance evaluation to become increasingly common and more complex in the near future; the cases that we discuss, thus, may be regarded as prototypes of future controversies. A key legal standard in the evaluation of teacher performance appears to be whether a teacher's students are demonstrating "normal educational progress." Although that notion may appear unequivocal, in reality such a standard opens up a plethora of types of data and data summaries of the student achievement scores that could be presented in legal proceedings. A recent important example of legislative concern with the evaluation of teacher performance using student test data is California Senate Bill 813, the Educational Reform Act; a court case in which these issues are prominent that we consider in detail in this report is Scheelhaase v. Woodbury, 349 F.Supp. 988 (1972) and 488 F.2d 237 (1974) from the state of Iowa.

Racial Discrimination: Arguments about the educational progress of a particular minority group serve as important evidence in unequal educational opportunity cases. Here the plaintiffs argue that the school district is not providing equal opportunity to all its students. The main indications of unequal educational opportunity arise in either comparison of the minority group progress to that of a majority group, or by the comparison of the progress of the minority group with some expected or "normal" educational progress. Specific examples of discrimination cases involving evidence of student progress are Hobson v. Hansen 269 F.Supp. 401 (1967) and 327 F.Supp. 844 (1971) in the District of Columbia; Serna v. Portales in New Mexico, 351 F.Supp. 1279 (1972) and 499 F.2d 1147 (1974); Johnson v. Sikes and Anderson v. Banks (aka, "Tattnall County") 520 F.Supp. 472 (1981), 540 F.Supp. 761 (1982), and 730 F.2d 644 (1984); and Berry v. Benton Harbor 515 F.Supp. 344 (1981).

Evaluation of School Desegregation Programs: As part of desegregation decrees or other legal agreements, the monitoring of student progress after the implementation of desegregation plays an important role. Concerns include the effects of desegregation on the progress of minority students and also any impact on the progress of the majority students. Two examples of monitoring of student progress subsequent to implementation of a desegregation program are Berry v. Benton Harbor (cited above) and San Francisco NAACP v. San Francisco Unified School District 576 F.Supp. 34 (1983). Both of these cases began as discrimination suits, which, after desegregation was ordered, involved subsequent monitoring of student progress.

Student Classification: A number of important court cases have involved assessments of student progress in ability grouping, tracking, and the assignment of students to special education classes. Several court cases involve assertions that a school district is practicing segregation under the guise of tracking. Claims of classroom racial identifiability and unequal educational opportunity for students have been the impetus for court action. A major method of sorting students is to group by ability either within a classroom or into different classrooms. Classification issues also surface for special class assignment. Public Law 94-142, the Education for all Handicapped Children Act of 1975, assures all handicapped children the right to a free and appropriate education. Numerous issues of educational progress are derived from this act, including the annual evaluation of student progress, both to re-evaluate the classification decision, and to comply with the requirement for federal funding. Individualized educational programs (IEP) provide one interesting method of monitoring the educational progress of handicapped children.

2.1 EVALUATION OF TEACHER PERFORMANCE

2.1.1 Scheelhaase v. Woodbury.

The appropriateness of the evaluation of teacher performance was the key issue in Scheelhaase v. Woodbury, 349 F.Supp. 988 (1972) and 488 F.2d 237 (1974), which was generated by the school district's failure to rehire a veteran teacher on the basis of her students' test scores. The teachers in the Woodbury School District are not tenured; rather, each year they are rehired at the district's discretion. Scheelhaase was one teacher who was not rehired; the school district claimed her performance was unacceptable. Scheelhaase contested the district's decision and brought the matter to District Court. The school district examined achievement scores of students who had completed a year in Scheelhaase's 7th grade classroom and concluded that the low scores reflected poor teaching. "The specific reason given plaintiff for termination was her professional incompetence, as indicated by the low scholastic accomplishment of her students on the Iowa Tests of Basic Skills (ITBS) and Iowa Tests of Educational Development (ITED)" (349 F.Supp. at 989).

In her complaint, Scheelhaase's expert witness, Norman Ashby, calculated the average gain per year of the entire school's students. These calculations were based on the average obtained grade equivalent scores on the subtests of the ITBS which was administered annually. The overall average student gains for each subtest were then treated as expected gains or standards. The gains of Scheelhaase's students were nearly the same as those averages, supporting the claim that her students made normal educational progress. (This analysis is considered in detail in Section 3.2)

Neither the court nor the defendant school district questioned these calculations. "The district court, McManus, Chief Judge, held that termination of Iowa teacher's contract by nonrenewal thereof on grounds of professional incompetence as indicated by low scholastic accomplishment of students on specified tests was arbitrary and capricious, since teacher's professional competency could not be determined solely on the basis of student's achievement on the tests, especially where the students maintained normal educational growth rates" (349 F.Supp. at 988-89). Thus, the district court held for the plaintiff, Scheelhaase.

The Appeals Court ruled for the school district. "The Court of Appeals, Talbot Smith, Senior District Judge, held that board's refusal to renew teacher's contract did not give rise to cause of action under Civil Rights Act" (488 F.2d at 238). "School board's decisions in exercise of its discretion are not vulnerable to correction by court merely if they are 'wrong,' sustainable only if they are 'right.' ...Such matters as competence of teachers, and standard of its measurement are not, without more, matters of constitutional dimensions which permit court to overrule school board's exercise of its discretion" (488 F.2d at 238).

This case provides one vivid example of the use of student gain or progress in evaluating teacher performance. In the Critique and Reanalysis Section (Section 3.2) we reexamine the evidence cited by both the District and Appeals Court.

2.1.2. California Senate Bill 813 (1983).

The Hughes-Hart Educational Reform Act of 1983 (SB 813) establishes a variety of programs designed to improve the quality of elementary and secondary education in California. A key area of this reform is teacher evaluation, and several sections of the bill address this issue. The obvious goal of educational reform in California is to improve the quality of classroom instruction. In this respect, then, all of the reform efforts eventually focus on the classroom. Teacher evaluation reform is only one piece of this larger reform effort designed to improve the quality of the student/teacher interaction.

SB 813 requires local districts to make the following changes: 1. The local board must adopt standards of expected pupil achievement by which teachers will be evaluated. The use of publisher's norms established by standardized tests are explicitly prohibited for this use. 2. Local districts must now identify the range of instructional strategies and techniques which teachers are expected to use appropriately. 3. Local districts must now ensure that curricular objectives are in place for all subject areas. Teachers will now be held accountable for the teaching of these objectives.

<u>Issues related to expected pupil achievement</u>. Although SB 813 mandates the districts to adopt standards of expected pupil

ERIC

achievement, little explicit guidance is provided. Certainly, making explicit standards for expected pupil achievement is a difficult and complex issue. Although a great deal of research has been done regarding effective schools and effective instructional practices, no elegant technical solutions emerge to the problem of linking teaching behaviors to student achievement measures.

In SB 813 the State has mandated that expected pupil achievement will be one of several criteria against which teachers in the state will be evaluated. Local districts are left to determine both the measures of student achievement and standards that will be used regarding this criteria. Indicators of achievement include tests, writing quality, or number of homework assignments completed. Standards, on the other hand, represent the level of a chosen indicator that is deemed acceptable. For example, one standard might be that 80% of a teacher's students should achieve a score of 70 or above on a district exam.

The use of student test scores alone as a standard of pupil achievement assumes a degree of comprehensiveness for such tests that they often do not possess. Exclusive reliance on objective paper and pencil tests may leave many important educational goals unexamined. An alternative might be the adoption of standards on a variety of measures of student achievement (homework, projects, essays, standardized tests, teacher constucted tests, and the quality of classroom discussions). (See Haertel, 1986)

Standardized tests matched to a district's curriculum do not exist in many subject areas. It may not be possible or wise for each district to spend time and money in the development of additional student tests to use in the evaluation of teachers. How to use existing measures in the construction of standards for expected student achievement is a complicated but important problem that SB 813 presents for local districts.

The educational reform movement in California and SB813 in particular, have attracted nationwide attention. The central role of expected student performance in the evaluation of teachers is likely to be a central issue of debate in any deliberation involving teacher dismissal or other action that is contested. The reason for including this legislative act in the review of court cases is its likely role in future court cases, along with its emphasis on the assessment of student progress.

2.2 EXAMPLES OF RACIAL DISCRIMINATION CASES

2.2.1. Hobson v. Hansen.

In Hobson v. Hansen, 269 F.Supp. 401 (1967) and 327 F.Supp. 844 (1971), the District of Columbia School System was found to "unconstitutionally deprive the district's Negro and poor public school children of their right to equal educational opportunity with the district's white and more affluent public school children" (269 F.Supp. at 406). The court ordered a desegregation remedy which consisted of a voluntary busing program. The background for the 1971 case was described by the court as: "In an effort to diminish the discrimination against children in the overcrowded schools east of the Park, this court in its 1967 decision ordered the Board to bus, on a voluntary basis, the primarily black and poor children from the overcrowded schools to the underpopulated, predominantly white nonpoor schools west of the Park. Several thousand children have been participating in the program during the past three years. Achievement test results from these children taken before they left the sending school compared with their annual test results at the receiving schools would have provided an indication, at least, of any discrepancy in the quality of the education available at the schools of one side of the Park vis-a-vis the other side" (327 F.Supp. at 858, footnote 26).

The most interesting and important aspect of this case from our standpoint is the court specification of what would constitute useful and compelling evidence of the adequacy of the

ERIC

district's desegregation actions to that date. "In an effort to suggest to defendants the kind of evidence they should be presenting if they were to prevail, the court ordered sua sponte on January 28, 1971 that 'defendants file in the record not later than February 15, 1971 such statistics and studies as will show the effect of the voluntary busing program on the achievement test scores of the children participating. These statistics should be on a school by school basis so that the improvement, if any, of the children at each receiving school may be discerned'" (327 F.Supp. at 858).

The court's idea in issuing this order was that a study of the improvement or lack of improvement in achievement test performance by students in the voluntary busing program, by providing a control for the factor of socioeconomic background, would be probative of whether schools west of the Park provide a better education than do schools in the rest of the city. On February 16, 1971, however, defendants moved the court to rescind this order on the ground that it imposes an unduly burdensome task or the defendants. The gist of the memorandum in support of the defendant's motion was that no systematic records of test results had been kept, and that those bused children who had been tested had been given different brands of tests for which conversion scales are unavailable--thus rendering meaningful comparisons impossible. "While the court does not charge the defendants with a lack of candor, it does seem incredible that a school system under injunction to provide equal educational

ERIC

opportunity to all its students would not have shown more interest in studying the effect upon individual student achievement of a voluntary busing program which permits students to transfer from allegedly inferior to allegedly superior schools. That defendants have failed to keep any systematic records of the achievement test results of these bused students raises questions about their effectiveness as administrators, if not about their good faith as parties to this case." (327 F.Supp. at 858-59).

What makes this case unusual is the clear emphasis on the effect that the busing program had on student progress as opposed to the less adequate but far more common reliance on a static assessment--how well do fourth grade students do when compared to national norms?--which does not serve nearly as well to reflect the effects of the desegregation program. To be specific on the nature of the usual static comparisons, consider this example of the plaintiff's evidence that unequal educational opportunity still existed in the district. "Plaintiff's prima facie case of discrimination in the provision of educational opportunity, based upon the pattern of unequal expenditures which favors the schools west of the Park, is strongly buttressed by further evidence in the record concerning the results of city-wide sixth grade reading achievement tests. The record shows that the west of the Park elementary schools produced an average reading achievement test score that was significantly higher--indeed 2.4 grades higher--than the average for the rest of the city. Obviously, these results tend to corroborate the presumption created by the pattern of expenditures that the city provide a better educational opportunity to its richer, white students. Defendants' dubious argument that the smaller classes and higher proportion of experienced teachers in the schools west of the Park do not give students there a better chance for a good education than can be had elsewhere in the city is still less convincing in light of this testing evidence." (327 F.Supp. at 858). Although this kind of static comparison is useful in identifying potential denial of equal educational opportunity, it is not adequate to assess the efficacy of an attempt to remediate. Inequities in test scores can be due to many other factors totally separate from the school district's provision of equal educational opportunity.

2.2.2 Serna v. Portales.

This case, 351 F.Supp. 1279 (1972) and 499 F.2d 1147 (1974), involved the Lindsey Elementary School in New Mexico and the education of Spanish-surnamed students. Plaintiff claimed that the school district was not supplying equal educational opportunity to the Hispanic students. The evidence presented consisted of both IQ and achievement data of the Lindsey School compared with the other three elementary schools (predominantly white) in the Portales School District. Lindsey School was the only school in the district with bilingual, bicultural programs. The evidence presented by the plaintiffs consisted primarily of the usual static comparisons of a grade level achievement of the minority school versus the other schools. The court found that "the evidence presented, indicates that the achievement of children at Lindsey is consistently lower than that of the children attending the other three elementary schools. I.Q. tests administered to fifth grade students in the four Portales municipal elementary schools reveal that the children at Lindsey scored approximately 13% lower than children at James and approximately 8% lower than children at Steiner . . . . What becomes apparent from an examination of these scores is that the performance of the children at every level at Lindsey School is not what it should be when compared to the performance of students at the other schools" (351 F.Supp. at 1281-82).

In the school district's appeal to the U.S. District Court, additional issues of cohort comparisons arise which suggest much better analyses regarding educational opportunity. "Undisputed evidence shows that Spanish-surnamed students do not reach the achievement levels attained by their Anglo counterparts. For example, achievement tests, which are given totally in the English language, disclose that students at Lindsey are almost a full grade behind children attending other schools in reading, language mechanics and language expressions. Intelligence quotient tests show that Lindsey students fall further behind as they move from first to the fifth grade. As the disparity in achievement levels increases between the Spanish-surnamed and Anglo students, so does the disparity in attendance and school

ERIC

dropout rates" (499 F.2d at 1149-50). The important point the District Court raises here is the notion that the static comparison at a single grade level does not tell the full or probably the important story; it is only by tracing the progress of cohorts of majority and minority students that one can find compelling evidence to adjudicate questions of "equal" performance or equal educational opportunity.

2.2.3 Tattnall County.

Tattnall County is how we refer to the series of cases (Anderson v. Banks and Johnson v. Sikes, 520 F.Supp. 472 (1981); 540 F.Supp. 761 (1982); 730 F.2d 644 (1984)) in which important issues about the use of test scores and student improvement in achievement entered into the Court's decisions. Here we give the background of the cases pertaining to discrimination issues. We will return to this case when examining classification issues (Section 3.4.1.2), and again in the Critique and Reanalysis section (Section 3.1), where we examine the statistical analyses of the test data presented to the court and describe our re-analyses and recommendations for data collection and analysis.

The Claim. The Tattnall County School District implemented the policy that high school students must be at at least the ninth grade level (GE=9.0) on the California Achievement Test (CAT) in reading and in mathematics in order to obtain a high school diploma. This policy was first instituted with the graduating class of 1978. The test has had vast differential racial impact: most of the students who did not pass were black.

Desegregation and implementation of the unitary school system in Tattnall County began in the 1970-71 school year. Before this time, the blacks in the segregated dual school system had substandard educational opportunities. The plaintiffs in this case felt that the diploma policy was racially discriminatory, given these past educational inequities.

Another relevant issue, following the case of Debra P. v. Turlington (474 F.Supp. 244, 644 F.2d 397), is when "the award of diploma depends on outcome of a test, burden is on school authorities to show that test covered only material actually taught" (644 F.2d at 475).

History. The Tattnall County schools were totally segregated until 1965. A voluntary integration system, encouraging blacks to attend the better white high schools was initiated, though it never got off the ground. 1970-71 was the first year the blacks and whites attended the same school. Despite the unitary system, no integration attempts were made: the buses carrying students to school were segregated, two buses for the blacks and two buses for the whites. There were no programs to help the students adjust. At the same time the unitary system was initiated, a tracking system was set up, which the school district claimed was grouping by achievement. Yet the tracking often resulted in racially identifiable classrooms with the higher tracks predominantly white, and the lower tracks predominantly black. Racially identifiable classrooms continued until the 1979-1980 school year, when the tracking system was abandoned as a result of an OCR investigation. (More on Tattnall's tracking system under classification issues.)

The test

The CAT was chosen by the school district without doing a local validation. The diploma requirement and establishment of remedial classes were passed by the school board in 1976, and were imposed two years later, beginning with the class of 1978. That class was administered the 197 CAT. Subsequent classes had the 1977 CAT forms C and D. CTB/McGraw-Hill developed the CAT. Grade-equivalency scores were derived from the norming sample. Steady progress through the school year was assumed in interpolating the grade equivalent scores for each month.

The plaintiff's expert witness, Dr. Shapiro, performed an item analysis on the CAT for Tattnall County using point-biserial correlations. He found that many of the items were biased in this population. In some instances the reliability of the CAT was quite unstable. It was assumed that progress from year to year should be steady, yet several instances were discussed where vast irregular fluctuations occurred from one year to the next. Yet overall, (steady) progress was made. The plaintiffs contend that it was the remediation program, and not the implementation of the exit exam, that was responsible for the improved scores.

The Court's decision. In order for the diploma sanction to violate the equal protection clause, it must be shown (1) that discriminatory racial impact occurred--which in this case is undisputed, and (2) that the policy was adopted with discriminatory purpose. The court did not find evidence of discriminatory purpose in Tattnall County's attempt to improve the value of a high school diploma. But the court also took into account the county's dual system history. "If present racially neutral actions serve to perpetuate the past intentional discrimination, there is no requirement that intent [(2) above] be proved again . . . . Thus, insofar as the poor performance of the black children is attributable to their participation in the dual system, the diploma sanction must fall. It cannot be constitutionally imposed on those who attended classes in the dual system" (520 F.Supp. at 500). "Clearly, no diploma sanction may be imposed until June, 1983, when those graduating will never have been exposed to the dual system" (520 F.Supp. at 503).

Will the diploma sanction be allowed after June of 1983? "Since the CAT is being used to measure what it was designed to measure, i.e., relative achievement levels in mathematics and reading, the court is of the opinion that local revalidation was not necessary" (520 F.Supp. at 507). Despite the examples of anomalous student changes (which will be discussed below) as an indication of the test's unreliability, in addition to the earlier discussion of the item bias study indicating an invalid test in Tattnall County, the "court cannot conclude that this

ERIC

well-respected test instrument is unsuitable for use in Tattnall County" (520 F.Supp. at 508).

But the curricula match is a different story. Drawing on Debra P., "the court can only conclude that where the award of a diploma depends on the outcome of a test, the burden is on the school authorities to show that the test covered only material actually taught . . . the court must conclude that the school authorities have not met the test of Debra P. . . . Because it has not been demonstrated that the items on the CAT were actually taught in the Tattnall County schools, the use of the CAT as an exit exam must fall on substantial due process grounds" (520 F.Supp. at 509). As Debra P.'s decision occurred simultaneously, the school district was initially unaware of this burden; thus, a new trial was granted for the sole purpose of displaying the curricular validity.

2.3 Examples of Evaluations of Mandated Desegregation Programs

Findings of discrimination are often followed by mandated remedies, either within the discrimination case or in a subsequent court action. Student progress, especially of the affected minority students, is a central concern in the success of the desegregation program in redressing previous discrimination.

2.3.1 Berry v. Benton Harbor.

Previous court opinions (442 F.Supp. 1280, 467 F.Supp. 630, and 494 F.Supp. 118) in 1977 and 1978 determined that the school district of Benton Harbor, Michigan was liable for "unconstitutional segregative conduct" throughout the school district through actions to "isolate black students in certain schools and within certain classrooms while preserving the predominantly white character of other schools and classrooms" (515 F.Supp. at 348).

The purpose of the present case is "establishment of a constitutionally acceptable, fair, understandable and workable remedy" (515 F.Supp. at 348). In framing the remedy for previous discrimination, "the court concluded that system-wide segregation such as the Benton Harbor District has created and the State Board of Education condoned and perpetrated by its inaction, result in measurably reduced achievement. . . . The problem of most concern to the court in addition to the racial separation and isolation of large numbers of children in these three school districts is the chronic low achievement levels of students in the Benton Harbor system. Dr. Stolee, the court's appointed expert, has recommended that its plan include a program to improve the quality of instruction provided to the students in the Benton Harbor School District and to raise the level of student achievement until achievement within the district reaches the average attained by Michigan students in the Michigan educational achievement program. Establishment of an achievement component in the Benton Harbor schools is crucial to any complete and effective remediation of harm that has resulted from the defendant's unconstitutional segregative conduct" (515 F.Supp. at 369).

A particularly interesting aspect of the court's remedy is the operationalization of the achievement goal. In describing the social skills and achievement component of the court's plan, the explicit statement is that "the goal . . . will be to raise the level of student achievement, until achievement within the district reaches the average attained by Michigan students in statewide achievement tests. Once that goal has been achieved, the program will be phased out over two additional school years" (515 F.Supp. at 370). Thus, this case provides an example of the monitoring of student progress as compared to a common norm or standard.

2.3.2 <u>San Francisco NAACP v. San Francisco Unified School District</u>.

ERIC

This case began as Johnson et al. v. San Francisco Unified School District et al. (339 F.Supp. 1315 (1971); 500 F.2d 349 (1974)), a school desegregation case in which it was found that "for a long period of time there has been and there now is de jure segregation in the San Francisco public elementary schools. This segregation must be eradicated forthwith" (339 F.Supp. at 1323). The district was ordered "to carry out, effective at the start of the next term of schools on September 8, 1971, desegregation of the student bodies of each and all of the schools as provided for by the Horseshoe Plan or by the Freedom Plan . . . To make bona fide, continuing and reasonable efforts during the next five years, to eliminate segregation in each and every school" (339 F.Supp. at 1324). Also the district was enjoined from a number of practices including "authorizing, permitting, or using tracking systems or other educational techniques or innovations without effective provisions to avoid segregation" (339 F.Supp. at 1325).

In 1983, the Northern California U.S. District Court considered a second case against the San Francisco Unified School District brought by the San Francisco NAACP (576 F.Supp. 34), in which a detailed consent decree was agreed to. Of particular interest is the portion of the consent decree that addresses academic excellence: "The decree directs the districts to monitor test scores and academic results in order to evaluate the continued effort to achieve academic excellence throughout the system" (576 F.Supp. at 42).

This case is unique in the emphasis given to the monitoring of student progress as a means to evaluate the efficacy of the desegregation plan. The 1983 court order specifies a consent decree which states in Paragraph 39 that "the S.F.U.S.D. shall evaluate student academic progress for the purpose of determining the curricula and programs most responsible for any improved test scores and learning in the District and the extent to which these curricula and programs are available to students of all racial/ethnic groups. The S.F.U.S.D. shall adopt any additional curricula and programs necessary to promote equal educational opportunity" (Consent Decree, p. 19). Paragraph 40 the Consent Decree requires an annual report which "shall include a section on S.F.U.S.D.'s progress toward the goal of academic excellence, setting forth test scores and other evaluative data for each building and for the District as a whole" (Consent Decree, p. 19). This case--the consent decree and the subsequent yearly reports--provides a special resource for the study of data on student progress and its use in judicial decisions.

The 1983-84 Annual Report pursuant to the Consent Decree by the San Francisco Unified School District (No. C-78 1445 WHO) reports the limited progress that the school district had made to that point. The 1983-84 Annual Report states that "The San Francisco Unified School District was directed to (1) determine curricula and programs most responsible for academic development" and "(4) provide test scores and other evaluative data for each school and the District as a whole" (C-78 1445 WHO, p. 15). At

ERIC

the time of this report the district had made limited progress. "The District had not fully met this requirement. The District did present information about curriculum programs contributing to academic development, the selection and/or adoption of such programs and an indication that no racial or ethnic restraints denied access to these programs" (C-78 1445 WHO, p. 15).

The most recent report, the 1985-86 Annual Report by the San Francisco Unified School District to the U. S. District Court, Northern District of California (No. C-78 1445 WHO) stated the first 1985-86 responsibility of the school district to satisfy Paragraphs 39-41 of the Consent Decree as "determine the curricula and programs most responsible for improved test scores and learning by evaluating student academic progress" (C-78 1445 WHO, p.1; emphasis added). The court found the school district to have fully satisfied this and other decree responsibilities. The court stated "during the 1985-86 school year, the San Francisco Unified School District using 1984-85 data has evaluated student academic progress for the purpose of determining the curricula and programs most responsible for any improved test scores and learning in the district and the extent to which these curricula and programs are available to students of all racial/ethnic groups." (1985-86 Annual Report, Vol. 4, p. 1).

The empirical evidence assembled by the San Francisco district sought to identify schools which have: "1. Consistently high relative and absolute student test performance and, 2. Schools where students rates of growth are faster than the national norm" (1985-86 Annual Report, Vol. 4, p. 2). The California Assessment Program scores were used for point 1. In particular, the district presented for each grade level tested by CAP (grades 3,6,8,12) the number of schools whose scores exceed the top of their comparison score band. The comparison score band is a statistical range constructed so as to include the middle 50% of the distribution of schools with demographic characteristics that are similar. Thus exceeding the comparison score band places a school's score in the top 25% of other statistically similar, in demographic terms, schools.

To identify schools with student growth rates faster than the national norm, the Comprehensive Tests of Basic Skills (CTBS) scores of fall of '84 and fall of '85 were used. The district calculated the "amount of growth experienced by students at each grade level at each school" by calculating the difference between grade equivalents scores of fall '85 and fall '84 testing, a simple difference score in the grade equivalent metric. Expected growth was defined by "national norm of ten months growth between annual testing." (1985-86 Annual Report, Vol. 4, p. 5). The district stated "Our strategy here was to identify grade cohorts within schools where the growth rate was greater than ten months. These were identified as grade levels where the programming was exemplary" (1985-86 Annual Report, Vol. 4, p. 5). The district found that in the '84-'85 data, 222 of the possible 342 elementary grade levels surpassed this national average. Furthermore the district reported, that the same analysis carried out on '83-'84 data revealed 195 of the possible 350 elementary grade levels surpassed the national average (i.e. the expected growth of 5/6 of a grade equivalent unit.) Furthermore, the district carried out the analysis of gains in grade equivalent units focusing on the original consent decree schools (see 1985-86 Annual Report, Vol. 4, p. 6). Analyses were also conducted on scores from those students who remained in their school for the full year (from fall '84 to fall '85). Achievement gains in terms of difference scores of grade equivalent scores were computed and found often to exceed the average rate of improvement of the district overall, even though the mean pretest and mean posttest scores for these schools were lower than that of the district average. These results were interpreted in a positive light as "any achievement rate greater than that of the district means that students were catching up to the district average" (1985-86 Annual Report, Vol. 4, p.7).

Section 2.3.3. San Diego City School district.

On December 2, 1980, the Superior Court of California, County of San Diego, issued to the Board of Education of the San Diego Unified School District a court order in the case of Kari Carlin et al. v. the Board of Education of the San Diego Unified School District (No. 303800, Order Re Integration Plan 1980-81). The court order stated in part the following: "It is ordered

ERIC

adjudged and decreed that: the Board of Education of the San Diego Unified School District will: 1. implement a course or courses of study in all minority isolated schools which will result, by the dates indicated in the table below, in 50% of the students in the isolated schools achieving at or above the national norm on the Comprehensive Tests of Basic Skills (CTBS) in reading, mathematics and language" (Re Integration Plan, p.1). The court's table started with kindergarten grades achieving this criterion by 1982, first and second grades by 1983, third through sixth grades by 1984, and seventh through eleventh by 1985.

In the fall of 1982 the court ordered additional analyses to provide a comparison of achievement levels in the minority isolated schools with that in the other schools in the district at all grades tested district-wide. In 1982 the court also requested the test data be provided by ethnic subgroups. On May twenty-first, 1985 the court issued its final order which incorporated all pertinent past orders including those relating to achievement scores. It ordered that the district make an annual report to the court by September first of each year in a format that would permit meaningful comparisons from year to year.

The explicit quantitative criteria set forth by the court make this case interesting, if not unique, as a case study in the use of test scores in desegregation cases. The district's reports give the percent of pupils scoring at or above the publisher's median for non-minority isolated schools and minority isolated schools for each of the major ethnic groups (Hispanic, white, black, Asian, Alaskan/Indian) for different years of testing. The most recent data are presented in Report number 425 of the San Diego City Schools Planning, Research, and Evaluation Division, "Testing Results from Minority Isolated Schools, Spring 1986" dated September 2, 1986. Figure 1 reproduces typical pages from Report 425. The longitudinal aspects of these data displays are quite interesting as the progress depicted is in terms of successive cohorts rather than of individual students improving over time. The trajectories are given to show how successive cohorts progress is constrained in order to meet the court mandated criteria of more than 50% of the students exceeding the publisher's norm. Although the court's criterion in any given year is a cross sectional one for each cohort, the role of individual student progress ma still be important. These data show unusual patterns of movement towards the court's criteria in the minority isolated schools, which may be due to cohort differences in improvement or to uneven progress by individual minority students as they move from one grade to the next. Such patterns of individual learning may severely confound the cohort by cohort comparisons featured in the district report.

- - - - - - - - - -

Insert Figure 1 here

- - - - - - - - - -

FIGURE SET 4B

GRAPHIC DISPLAY OF STUDENT PROGRESS

PERCENT OF PUPILS SCORING AT OR ABOVE THE PUBLISHER'S MEDIAN

MATH - FORM U DATA

JUNIOR HIGH SCHOOLS

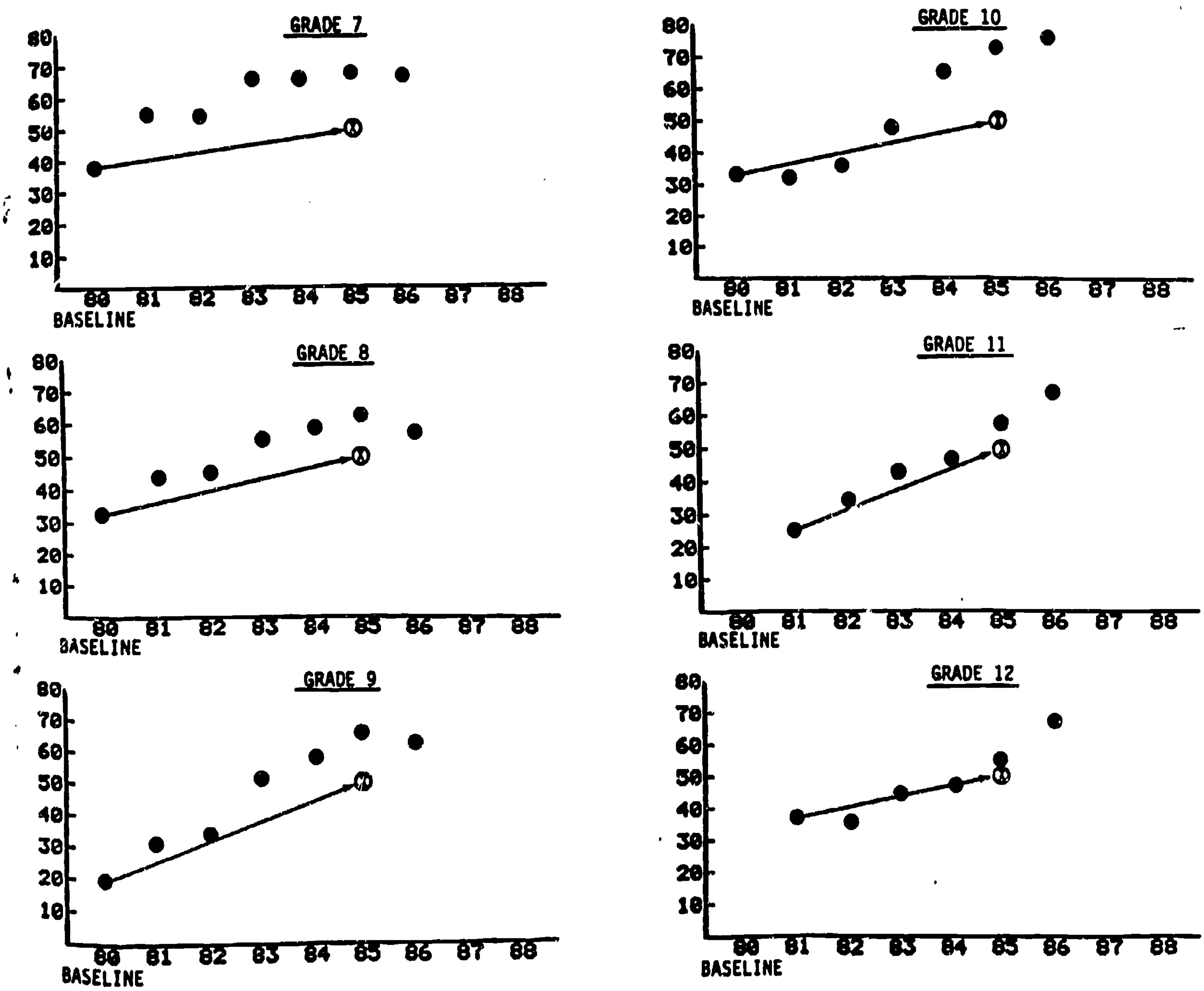
GRADE	1980 (BASELINE)	81	82	83	84	85	86	87	88
7	38.3	53.3	53.2	66.4	66.4	67.6	66.9		
8	31.0	42.7	43.2	55.3	59.4	61.6	56.7		
9	19.8	30.3	32.9	50.8	58.2	65.5	62.4		

PERCENT OF STUDENTS - PUBLISHER'S MEDIAN

SENIOR HIGH SCHOOLS

GRADE	1980 (BASELINE)	81	82	83	84	85	86	87	88
10	32.3	31.8	35.8	48.2	67.1	71.6	75.9		
11	X	25.9	34.9	44.1	47.2	58.4	67.8		
12	X	38.6	36.4	44.6	47.5	55.0	68.0		

PERCENT OF STUDENTS - PUBLISHER'S MEDIAN



⊗ GOAL FOR THE RESPECTIVE GRADE LEVEL AS EXPLICATED IN COURT ORDER (i.e., YEAR FOR ATTAINMENT OF PUBLISHER'S MEDIAN)

27

Figures 1a and 1b. Descriptions of progress of successive cohorts towards the court;s criterion presented by San Diego district.

Figure 1b.

GRAPHIC DISPLAY OF STUDENT PROGRESS

PERCENT OF PUPILS SCORING AT OR ABOVE THE PUBLISHER'S MEDIAN

GRADE LEVEL: 7

CONTENT AREA: TOTAL READING

NORMS USED: FORM U

GROUP / DATA		FALL 1979 MI	FALL 1979 NON-MI	SPRING 1986 MI	SPRING 1986 NON-MI
HISPANIC	N	100	837	150	1010
	%	12	36	15	36
WHITE	N	49	4049	147	2682
	%	73	70	85	74
BLACK	N	165	716	185	778
	%	12	30	32	44
ASIAN	N	9	499	30	1119
	%	22	56	57	53
ALSKN/INDN	N	0	19	0	18
	%	--	68	--	44
COMBINED	N	323	6120	512	5607
	%	21.7	60.0	43.8	58.6

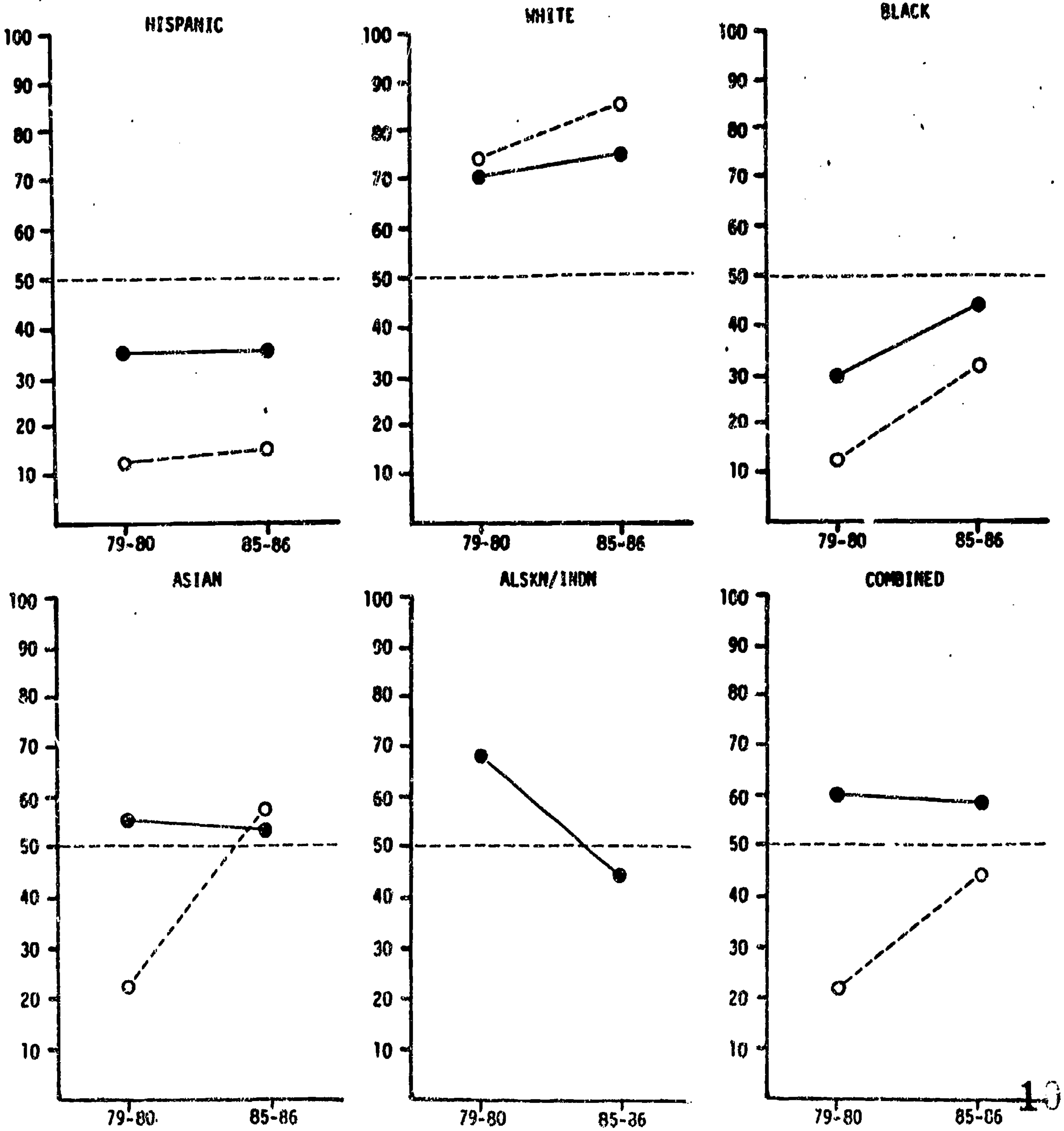
N = NUMBER OF STUDENTS

% = PERCENT SCORING AT OR ABOVE THE PUBLISHER'S MEDIAN

CODE:

●———● NON-MINORITY ISOLATED SCHOOLS

o- - - -o MINORITY ISOLATED SCHOOLS



59

Section 2.3

ERIC

2.4 Examples of Student Classification Cases

Our heading "student classification" encompasses two distinct kinds of topics of litigation: Ability grouping of students within classrooms and within schools, and assignment to special education classrooms.

2.4.1 Ability Grouping or Tracking.

Key issues relating to student progress in schools involve the efficacy or lack thereof of tracking. In some cases, the claims of racial identifiability and unequal educational opportunity for students in lower tracks have been reasons for court action. Other litigation involving ability grouping is separate from the termination of racial discrimination. Examples of court cases which involve assertions that a school district practices follow the general discussion of ability grouping.

One method of tracking is to group students by ability. "Ability grouping is the practice of organizing classroom groups in a graded school to put together children of a given age and grade who have most nearly the same standing or measures or judgments of learning achievement or capability" (Findley and Bryan, 1975, p. 9).

Findley and Bryan (1975) posit 35 conclusions regarding the practice of ability grouping. Their discussion draws from the results of the U.S. Office of Education Task Force, in which Findley was the principal investigator. The conclusions regarding monitoring student progress follow.

"Much reliance is placed on early determinations of capability. Individual children are often not reevaluated for different group placement . . . . Initial differences are thus allowed to cumulate so that low groups fall progressively further behind" (p. 21).

Findley and Bryan urge that "provision should be made for frequent review of each individual's grouping status as part of the instructional program. The evidence that ability grouping results in practically permanent assignment of children to low or high groups . . . makes a regular program for reviewing group placement absolutely essential" (p. 27). Inherent in the plea for mobility between tracks is some method for measuring an individual's progress.

Another conclusion reads that the practice of ability grouping produces "almost uniformly unfavorable evidence for promoting scholastic achievement in average --or low-- achieving groups" (p. 13). A description of Findley's method for the analysis of scholastic achievement was given in Section 1.3. An example in that discussion displayed mostly less than expected progress for the low ability levels. Not all the studies assessing the change of lower track students show such losses. In view of the conflicting results, researchers "agree that the evidence cannot be used to support the assumption that ability grouping aids achievement. Furthermore, these reviewers conclude that where research studies cite achievement gains in favor of homogeneous [ability] grouping, results are often explainable on the basis of differing teaching methods and materials, modification of educational objectives, and curriculum reorganization" (Wilson and Schmits, 1978, p. 536).

Oakes (1985) surveyed secondary teachers regarding their grouping practices. Shes concludes that ability grouping persists not only out of tradition, but because teachers believe that students profit most if they are with those "who learn at the same rate" (p. 6). Oakes' sentiments seem to be representative of current educational researchers. "As we know from the research . . . tracking . . . does not appear to be related to either increasing academic achievement or promoting positive attitudes and behaviors" (Oakes, p. 191).

All the researchers reviewed believe that the best grouping practice consists of small groups of individuals with heterogeneous abilities, sometimes referred to as the "Baltimore plan." If ability grouping must persist, then mobility between tracks, and thus careful monitoring of progress, must prevail.

2.4.1.1 NAACP v. State of Georgia.

This case, NAACP v. State of Georgia (CV482-233 U.S. District Court, Georgia) involves the contention that ability grouping is an attempt to disguise intentional racial discrimination. In the ruling that certified class action status for the plaintiffs' suit, (99 F.R.D. 16 (1983)), the complaint was summarized as: "The plaintiffs seek declaratory and injunctive relief against the defendants in order to end alleged intentional racial discrimination in the public schools in the state of Georgia. Plaintiffs contend that black school children in Georgia, including the individual plaintiffs and the members and children of the state conference's branches in Liberty County NAACP have been assigned to classrooms within biracial schools on the basis of their race. The plaintiffs allege that the assignment of these children to disproportionately black classes results from the operation of intentional racial discrimination rather than racially neutral criterion such as ability or achievement grouping. Plaintiffs claim that as the result of the disproportionate classroom configurations they and the class they seek to represent have suffered severe educational deficits which must be remedied" (99 F.R.D. at 19).

The case also contains a special education component which is summarized as: "Plaintiffs also seek relief from allegedly racially discriminatory administration of the special education program in the state of Georgia. It is alleged that black school children are erroneously classified as being educable mentally retarded as a means of removing them from normal classrooms or excluding them from programs for specific learning disabled children. Plaintiffs further contend that this misclassification of black children is racially motivated and has a devastating effect on these children's lives because of the stigma associated with being identified as mentally retarded and the diminution of their educational opportunities. Finally, plaintiffs claim that the misclassification of non-handicapped or specific learning disabled children as educable mentally retarded also excludes them from appropriate academic placements because of a real or attributed handicap" (99 F.R.D. at 19).

In 99 F.R.D. 16, the district court certified class status for the plaintiffs and the case was then tried in the U.S. District Court for the Southern District of Georgia, Savannah Division. The two basic claims of the case are "(1) that black children have been tracked into racially malapportioned classrooms in otherwise integrated schools; and (2) that black students have been misclassified as being Educable Mentally Retarded (EMR) because of their race" (99 F.R.D. at 19-20).

The court ruled for the defendant school districts on both claims. The court found no intentional discrimination on the basis of race or handicap. Some of the evidence supporting the defendant school district's case (to the dismay of the plaintiffs) included student gains. "The local defendants demonstrated that significant academic gains have been made by students educated within the structured learning environments. Effectiveness of the ability grouping models for children instructed in the low achievement groups was particularly emphasized" (Civil Case Judgment Marshall v. State of Georgia CV482-233, p. 22).

It is of particular interest in this research to examine the defendant's evidence of student progress. Some of this evidence, excerpted from pages 22-24 of CV482-233, follows.

> "The students participating in Chapter 1 programs in the Vidalia City schools have consistently exceeded the gains expected pursuant to the approved Chapter 1 program. In 1978, when the expected gain in number

ERIC

systems, computation, and application was 1 Normative Curve Equivalent (NCE), only three students from among over 200 tested, failed to meet this expected gain. The remaining students experienced NCE gains ranging from 1.8 to 6.1. Likewise, in 1980, the actual NCE gain experienced ranged from 2.4 to 19.9 NCE's. Similar progress was shown for 1982. Vidalia also illustrated its improvement in performance on the GCRT in math and reading. An analysis of test results for the Spring of 1981 and 1983 demonstrates an average increase in objectives achieved out of a possible total of 20 of 4.04 by Chapter 1 students while non-Chapter 1 students experienced an increase of average objectives of only 1.27.

"Lee County showed similar gains for its Chapter 1 students. In the 1980-81 school year, the average improvement shown by these students, in grade equivalence, was 1.45 grades in reading and 1.90 grades in math. In the next term, the average improvement was .88 grade equivalence in reading and 1.38 in math. Effectiveness of Lee County's educational methods for all students was demonstrated by the students' achievement on the GCRT. For example, in 1983 the second graders exceeded the Georgia state average in 16 of 20 math objectives measured, equalled the state average in 3 of those objectives, and fell below in only 1. In the same year, the second graders' reading scores exceeded the state average in 24 of 25 objectives measured and equalled the state average in the 25th. Similar results were experienced by Lee County's third and fourth graders.

"Crisp County's students have also exhibited academic progress since the introduction of achievement grouping. This is seen in GCRT results between 1976 and 1980, during which time the percentage of fourth graders who satisfied GCRT reading and math objectives rose from 59.2% to 80% and from 59.1% to 77.5%, respectively. For the same period, the percentage of eighth graders who satisfied the GCRT objectives increased 26.2% for reading and 19.4% for math.

"Criterion Reference Test tables for Evans County reveal that its students have also made significant academic gains. Over the past five years, Evans County children have outscored their counterparts across the First District in 82.5% to 97.5% of the test objectives and outperformed the state as a whole in 60.0% to 92.5% of the GCRT objectives."

The plaintiffs presented a number of statistical analyses using student test scores which they argued showed harmful educational impact of the ability grouping. This evidence was not found to be persuasive in the judgment. (See pages 30-38.)

2.4.1.2 Grouping Practices in Tattnall County Case.

Analysis of Tracking System. The court found the disproportionate number of blacks in the lower tracks difficult to justify. Based on an available sample of IQ scores, the court concluded that "the black and white children who were later subjected to the CAT diploma requirement began their academic careers with generally equal abilities" (520 F.Supp. at 481). The school district claimed that track placement was based on achievement, not ability. The following were used as criteria for track assignment: Scott-Foresman Initial Placement Score, teacher judgment, and CAT scores. The children are grouped together for all academic subjects. The lower groups were taught essentially the same material as were the higher groups: No remediation was offered.

Collins High was the one Tattnall County school that did not practice the tracking system. These students served as comparisons for the performance of those students who were tracked. Two standardized tests served as the basis for comparison: The Georgia Criterion Referenced Test (GCRT) and the California Achievement Test (CAT). For eighth graders in 1978 and 1979, the students from the non-tracked school were dramatically ahead on both the GCRT and CAT. The principal of Tattnall Elementary School submitted evidence that compared "pass rates" (i.e., proportion of students exceeding 9.0 on the CAT) of high school students from the tracked and nontracked high schools. For the years 1978, 1979, and 1980, a greater proportion of tracked black students passed than did non-tracked blacks (520 F.Supp. at 484). Following this conflicting evidence, the judge wrote "I find it impossible to accurately evaluate the effects of achievement grouping" (520 F.Supp. at 485).

2.4.1.3 Dillon County

The Office of Civil Rights, United States Department of Education (OCR) initiated action against the Dillon County School District Lakeview, South Carolina alleging that the district was in violation of Title VI of the Civil Rights Act of 1964 because of the district's failure to alter or abandon its practice of assigning students in grades 1-8 to racially identifiable ability grouped classes, without educational justification. The OCR further alleged that compliance by the district with Title VI could not be achieved by voluntary means and thus sought an order of relief terminating and refusing to grant or continue further federal financial assistance to the district until the district could satisfy the OCR that it had appropriately remedied its alleged failure to assign students on the basis of educationally valid, non-discriminatory procedures and that it would comply in the future with all applicable requirements of Title VI of the Civil Rights Act of 1964. The district denied the allegations of the OCR and claimed that its student assignment methods are in no way motivated by racially discriminatory intent but rather by the purpose of providing better educational opportunities. The failure of OCR and the district to reach a voluntary agreement led to an administrative hearing, in the U.S. Department of Education, (docket number 84-VI-16, compliance proceeding under Title VI of the Civil Rights Act of 1964, Eugene Powell, Jr. presiding judge).

Starting in school year 1982-83, the district began assigning students to regular ability-grouped classes in grades 1-8 by relying primarily on the functional equivalent of a rank ordering of the Comprehensive Test of Basic Skills (CTBS) scores. The district employs the "block placement" approach in grades 1-8 with the CTBS total battery scores a primary criterion for placement. Students in grades 1-6 are placed by the block method within self-contained ability levels for the entire day for all subject areas with the exception of those students who are pulled out of the classroom for Chapter 1 remediation in either reading or mathematics. Students in grades 7 and 8 are assigned to departmentalized classes whereby they move together as a block from subject to subject for the entire academic day. The kindergarten and high school placement methods are exceptions to the block placement method of assignment.

The result of this ability grouping on the basis of CTBS total score is a large number of "racially identifiable" classrooms; for the school year 1983-84, 19 out of the 34 classes in kindergarten through eighth grade were racially identifiable. The criteria for racial identifiability is the 20% "rule of thumb" or "trigger" used for many years in discrimination cases by many government agencies as an initial presumption of adverse discriminatory impact. The class is judged "racially identifiable" when the black percentage composition of a class differs by more than twenty percentage points from the overall or average black percentage composition of the grade as a whole. Moreover, the top ability levels always contain the lowest percentage of blacks, and the lower ability levels contain the highest percentage of blacks. These results were presented as highly statistically significant and practically important.

Two important types of questions about student academic progress were prominent in the hearing: 1. Student mobility between tracks over time and, 2. The academic progress over time of students in the lower track. Mobility between tracks can be examined by the analysis of transition matrices for adjacent years, such as the transition matrix Figure 1, or more comprehensively by multivariate versions of the transition matrix for more than two years. Of particular importance is whether the transition matrices (and therefore specific transition probabilities) differ for white and black students. An analysis of mobility between tracks presented in the administrative hearing revealed a "pattern that black students tend to move down in the ability groups and white students tend to move up. If the white student does move down in ability level, he or she tends to bounce right back up. There is only a 17% likelihood that whites will move down from the top track while there is a 39% percent likelihood that they will move up from the bottom track. The pattern is exactly reversed for blacks in that 31% of the time blacks moved down when they had reached the top level and 21% of the time they move up from the bottom track. There is a similar pattern in analyzing the middle track although this pattern is not as dramatic as the analysis of the top and bottom tracks" (pp. 53-54 of Administrative Proceedings). Differential mobility among ability groups or tracks is an important general issue, and one which needs considerable methodological and statistical attention.

- - - - - - - - - - -

Insert Figure 1 here

- - - - - - - - - - -

A second issue is the effect of the ability grouping on student progress. An analysis presented by OCR centered on students with the same or matched subtest score (e.g. math) but who were in different tracks (track placement was determined by the overall battery score). Examining those pairs of students' subsequent achievement scores for that subtest indicates whether there is differential progress in that subtest area. A discernable pattern was found in the relationship between placement in a track and subsequent academic progress and performance. In some, being placed in a relatively high track leads to relatively better performance while being placed in a relatively low track would seem to lead to relatively poor performance. The ability group level in which a student is placed, even though he or she begins with the same subtest score,

PARKVIEW MIDDLE SCHOOL

MOBILITY BETWEEN GRADE 7 AND 8

EIGHTH GRADE TRACK	1	2	3	4	
4	2 (B=2)		5 (W=1, B=4)	13 (W=4, B=9)	n=20
3		3 (B=3)	13 (W=6, B=7)	2 (B=2)	n=18
2	6 (W=6)	15 (B=4, W=11)	3 (W=3)		n=24
1	26 (B=6, W=20)	3 (W=3)			n=29

SEVENTH GRADE TRACK

Figure 1. Transition matrix showing mobility between seventh and eighth grade ability groupings (1 high, 4 low) for Black (B) and White (W) students.

directly correlates with whether he or she will do better or worse on a test the following year. Being placed in a low track is to the student's detriment compared to placement in a high track. These findings have statistical and practical significance and show that although students may start out with roughly comparable scores, by the end of the year, the placement in a different track has increased the gap between them (pages 54-55 of the Administrative Hearing). The district attempted to use evidence of academic progress to find an educational justification for the ability grouping; favorable statewide testing results and a 2% annual increase (in percentile rank metric) among the lower groups in almost all grades were part of the district's defense of the educational justification of its ability grouping.

Another important issue related to ability grouping that was prominent in this case is the superiority of bivariate or multivariate classification as opposed to univariate classification. By univariate classification we mean practices similar to the Dillon County District, where a single score, composed of the aggregated math, reading, and language subtest scores, is used to assign a student to a certain ability group and that student assignment stays the same in all the relevant subjects. A multivariate classification would use the relevant subtest score to group students for that domain; e.g., students are grouped by their mathematics ability (based on the mathematics subtest scores) for their mathematics classes, by

ERIC

their reading ability for reading classes. This multivariate classification would, for example, allow a student to be in the high group on mathematics and perhaps in the lower group on reading. Testimony regarding the bivariate classification for math and reading led the judge to conclude that ability grouping done for math and reading on the basis of the respective subtest score with the remainder of the subjects heterogeneously grouped by random method would result in approximately 32% (or using the time weighted theory, 23%) of the classes being racially identifiable (p. 74 of the Administrative Proceeding). This is far superior than the 52% of racially identifiable classes with present univariate grouping.

As a result of the evidence presented on racial identifiability resulting from ability grouping, the differential mobility between blacks and whites, and differential academic progress and achievement as a result of ability grouping the administrative law judge concluded that Dillon County is in violation of Title VI of the Civil Rights Act and that the U. S. Department of Education has authority to terminate all federal financial assistance under Chapter 1 and Chapter 2 of the Educational Consolidation and Improvement Act of 1981.

2.4.2 Legislation and Litigation Regarding Special Education

Progress of special education for handicapped students in alternative school settings has been prominent in many judicial decisions. The Education for All Handicapped Children Act of 1975 (Public Law 94-142, 89 Stat. 773) assures all handicapped children the right to a free and appropriate education. "It is the purpose of this Act to assure that all handicapped children have available to them . . . a free appropriate public education which emphasizes special education and related services designed to meet their unique needs . . . and to assess and assure the effectiveness of efforts to educate handicapped children" (89 Stat. at 775). One way to assess the effectiveness of efforts to educate handicapped children is to provide an individualized educational program (IEP) for each child. "The term 'individualized education program' means a written statement for each handicapped child" which shall include at least the following: "a statement of the present levels of educational performance of such child, a statement of annual goals, . . . appropriate objective criteria and evaluation procedures and schedules for determining, on at least an annual basis, whether instructional objectives are being achieved" (89 Stat. at 776). The IEP must include initial status, the instructional goals, and a method to monitor the student's progress toward meeting these goals. These IEPs are to be revised at least annually: once a year, progress is to be assessed and new objectives outlined. The details of the use and formulation of IEPs has been discussed in the section on data and methods for assessing student progress.

2.4.2.1. Special Education Issues in Tattnall County Case

Tattnall County places a larger proportion of its students into special classes than does the state of Georgia or the nation. "Placement in the EMR program for a student who is not truly retarded is disastrous . . . . The longer the child is misclassified, the more severe and difficult to repair the damage caused by the misclassification becomes" (520 F.Supp. at 495).

The state of Georgia regulates the procedures for placement into special classes. A student is first referred for evaluation, then is given a physical examination, and then psychological tests. A Special Education Placement Committee is to decide on the appropriate program for the child. An individual educational program is drafted and revised annually. The placement decision is to be reevaluated every three years. Parental consent is necessary before the student may be placed in a special program.

In Tattnall County, a "grossly disproportionate number of black children were classified as mentally retarded while the classes for children with specific learning disabilities were mostly white" (520 F.Supp. at 496). The disproportionate racial impact implied that some children were misplaced. Examples were provided of several students who were placed in EMR classes yet had IQ or CAT scores well above the cutoff for EMR placement. It was discussed that average students are expected to gain about a year in achievement (presumably, a point in grade-equivalence scores) for each chronological year, whereas slower students are expected to gain at slower rates (520 F.Supp. at 497-98).

In regard to the evident misclassification of students into special classes, the Court requested that the plaintiffs and defendants submit a plan to the court to provide remediation for misclassified students and a plan for a reevaluation of all students classifed as EMR.

2.4.2.2. <u>Hoffman v. Board of Education of the City of New York</u>

The background for this educational malpractice case 410 N.Y.S.2d 99 (1978), was that Hoffman scored 74 on an IQ test in first grade and was placed in a class for "Children with Retarded Mental Development (CMRD). He had a speech impediment (since his father's death) and was taken to a speech disorder clinic at about the same time. At the time of the IQ testing, the school psychologist recommended that Hoffman be "reevaluated" within two years, since his IQ score was virtually on the cutoff for special class assignment. Hoffman remained in CRMD classes until the 12th grade, at which time all those students were transferred to an Occupational Center for the handicapped. To verify eligibility for this occupational program (i.e., mentally handicapped), all the students were administered IQ tests. Hoffman scored in the 90s--well within "normal" range. He was therefore disqualified from attending this vocational school.

Hoffman sued the school district for educational malpractice. His claim included digging up the score on a non-verbal IQ test given to him at the Speech Disorder Clinic twelve years prior. He had scored in the 90s then, and claimed that he should not have been retained in the special classes for his entire school career. Rather, the school district had the obligation to reevaluate him within the two years after initial placement, as the school psychologist had recommended. To the plaintiffs, "reevaluate" meant "retest."

The defense claimed that "reevaluate" meant just that--and that all students are reevaluated continuously by their teachers. In addition, district-wide achievement tests are administered twice yearly, and Hoffman's scores were "still well below what would be expected of a child in a regular class" (410 N.Y.S.2d at 103). The defendant claimed that, had the boy's scores been higher, he would have been retested on ability and possibly placed into regular classes. They did not contest the boy's "true" IQ of 96.

Hoffman prevailed in the District Court, but the decision was reversed in appeals court. The appeals court did state that, "Fortunately, since 1968 no such error is likely to occur for both New York State and the New York City Board of Education now require frequent and periodic intelligence retesting of children in CRMD classes" (410 N.Y.S.2d at 108).

The issue of general concern is that Hoffman was a child of normal intelligence who was erroneously placed in the CRMD classes. Could we expect him to have high enough achievement scores (on the semi-annual tests) to signal the administrators to retest (IQ) him? How much is a child of normal intelligence expected to learn in a restricted learning environment? How can this progress properly be assessed?

SECTION 3

CRITIQUE AND REANALYSIS

The reanalyses in this section illustrate certain types of statistical analyses of student progress, which are based on Rogosa's prior technical research. Detailed attention is given to the special issues arising with data from educational court cases. The two sets of data we were able to obtain for this project were from the Tattnall County case (sections 2.2.3 2.4.1.2 and 2.4.2.1) and from Scheelhaase v. Woodbury (section 2.1.1). The reanalyses allow both a critique of the kinds of data and data analyses used in the court cases and also demonstrations of some improvements. In section 4 the details of improved analyses of student progress are further explained.

ERIC

3.1 TATTNALL COUNTY DATA REANALYSIS

Access to the raw data upon which some of the plaintiff's testimony was based allowed detailed examination of the quality of the data in the court case and our independent reanalyses. We describe the available data, our descriptive analyses of student progress, and statistical estimation of curricular correlates of student progress. Particular attention is given to the metric (and internal consistency) of the achievement test data including a comparison of statistical analyses based on grade equivalents and raw scores. These data are of particular interest because of the prominent role of questions about student progress in the testimony and the court's decision.

3.1.1 The Data

The data which are used in these analyses were supplied to Dr. Robert Calfee (Stanford University, School of Education) during the course of his involvement with the Tattnall County cases (sections 2.2.3 and 2.4.1.2). The data consist of scores and background information for a cohort of Tattnall County high school students who were freshmen in 1976 and who were tested yearly until graduation in 1979.

The data analysis files come from two sources: (1) a xerox listing of reading and mathematics raw scores and grade equivalents (GE's) for 212 students, and (2) a computer file (called GAREC) created for Dr. Calfee's analyses of a sample of 50 students. The latter file also contains background data on each of the 50 students as well as numerous variables created through Dr. Calfee's analyses. Calfee's analyses are described and evaluated by the court in Anderson v. Banks 540 F.Supp. 761 (1982, pp. 762-5).

A three-digit identification number existed for each student on the xerox listing (n = 212) and codes were also present for "race," "gender" and "school attended." Gender was used as a background variable in early analyses, before GAREC was located. From the variables on GAREC, four were selected as most promising for subsequent work. These were ELEMREAD, ELEMMATH, FINEENG and FINEMATH. ELEMREAD and ELEMMATH are identified as "GR 4-8 GPA [average grade point average in the subject over grades 4 through 8]" for reading and mathematics, respectively. They are coded as whole numbers, 1 through 4. FINEENG is a "fine curriculum index" in english and FINEMATH is a similar index for mathematics. From Dr. Calfee, we were able to ascertain that they are coded so that higher values reflect more exposure to regular courses, and lower values correspond to more remedial courses.

It will be noted that only the grade equivalent scores appear in GAREC. However, we were able to associate background variables with raw scores by matching student identification numbers and merging the appropriate information from hard copy and tape files.

The California Achievement Tests (CAT) were used for the testing in Tattnall County (see discussion of CAT in Anderson v. Banks and Johnson v. Sikes 520 F.Supp. 472 (1981, pp.485-94). However, the precise edition, level and form were not specified in any of the historical records made available to us. This became important as some of the data were incomplete and apparent inconsistencies existed in those data that were present (see also the court's discussion 520 F.Supp. at 492).

After considerable investigation and analysis it appears that the 1970 edition of CAT, Level 5 Form A (CAT/5A) (and possibly Form B as well, though this has not be verified) was used for the 1976, 1977 and 1978 testings. The 1977 edition of CAT, Level 19 Form D (CAT/19D) appears to have been used for the 1979 testing of the students as 12th graders. Table 1 shows the maximum raw score, the maximum GE, and the raw score at which the maximum GE is first attained for each of the tests.

Insert Table 1 here

Descriptive analyses. The full data set consisted of records for 212 individuals. However, not all individuals had complete records of test scores at all four administrations. Table 2 shows means and standard deviations of raw scores and grade equivalents for those students who had complete testing information in reading and in mathematics. (Inexplicably, some students had GE's when no raw score appeared on the file. Consequently, there are more students with complete GE records.) Note that the GE's increase each year in magnitude and maintain roughly the same variance. However, the raw scores in both reading and mathematics show a drop in Grade 12 (recall that there are fewer items on CAT/19D than on CAT/5A).

Insert Table 2 here

3.1.2 Initial Statistical Estimation of Growth

For the first set of analyses, we take the data simply "as

TABLE 1
CHARACTERISTICS OF THE TESTS

TEST	Max Raw Score	Max GE	Raw Score at which Max GE is attained
CAT/5A Reading Total	85	13.6	67
CAT/5A Mathematics Total	98	13.6	77
CAT/19D Reading Total	70	12.9	46
CAT/19D Mathematics Total	85	12.5	53

TABLE 2
MEANS AND STANDARD DEVIATIONS
TATTNALL COUNTY FULL DATA

READING:

GRADE	RAW SCORE		GRADE EQUIVALENT	
	Mean (n = 147)	S.D.	Mean (n = 152)	S.D.
9	36.9	12.7	7.7	2.6
10	39.5	14.5	8.6	2.9
11	42.2	14.7	9.1	2.8
12	41.3	12.8	10.0	2.1

MATHEMATICS:

GRADE	RAW SCORE		GRADE EQUIVALENT	
	Mean (n = 146)	S.D.	Mean (n = 152)	S.D.
9	42.3	18.4	7.8	2.6
10	50.5	19.6	9.1	2.8
11	56.7	17.7	10.0	2.6
12	50.2	14.3	10.7	1.9

is" except for arbitrarily supplying a constant value, zero, as the decimal value in the grade 12 reading GE. (Apparently a similar approach was taken with the analyses presented to the court, since the 12th grade reading GE's in GAREC all ended with ".4".) Then individual growth curves are fit to each student's data. Analyses are done separately for Reading Total raw scores, Reading Total GE's, Mathematics Total raw scores and Mathematics Total GE's.

Since only four points in time are available (the testings in grades 9, 10, 11 and 12), straight-line growth curves are fit using ordinary least squares. Thus, the model for person p is

$$X_p(t) = \xi_p(0) + \theta_p t + \varepsilon$$

where

$X_p(t)$ is the observed score for person p at time t ,

$\xi_p(t)$ is the true score for person p at time t ,

θ_p is the true linear growth rate of person p ,

t is time and

ε . is random error, assumed to be independent at each time.

Fitting this model allows the estimation of a constant rate of change for each student. We also examine the R^2 values for each fit to get some idea of how adequate a straight-line model is for these data. Five number summaries of $\hat{\theta}$ and R^2 are shown for each of the four analyses in Table 3. It appears that grade equivalent scores give better results than raw scores. Note the median R^2 is 60.4 and 63.3 for reading and math GE's, respectively; compare this with 33.6 and 25.1 for reading and mathematics raw scores, respectively. Also note that the median rate of change is slightly less than one GE per year in both reading and mathematics. Statistical proocedures for this estimation have been implemented for this project in SAS for mainframe computers and in the language GAUSS for MS-DOS microcomputers; an appendix to this report contains the listings of these programs.

Insert Table 3 here

Table 4 shows estimates, computed from the Tattnall County data, of a number of these quantities which describe properties of a collection of individual growth curves.. To briefly summarize, t° and κ characterize the time scale; σ^2_{θ} amd $\sigma^2_{\xi(t^{\circ})}$ describe the variance in growth rate and status; $\rho(\hat{\theta})$ and the standard error of $\hat{\theta}$ characterize the reliability and precision of the estimates of straight-line growth rate; γ is an index of tracking that describes the stability of individual differences over time; and $\rho_{\theta W}$ is the correlation between growth rate and some background variable, W . For more detailed descriptions of these quantities, see Rogosa and Willett (1985).

Insert Table 4 here

Tables 5 and 6 give some indication of the consistency of high and low scoring students. The student identification numbers are listed for those students with the five lowest and five highest values of the variables shown. For example, note that student #109 has the second highest value of R^2 among the fits for Reading Total raw score and the third highest on Mathematics Total raw score. Not surprisingly, this student has

TABLE 3
FIVE NUMBER SUMMARIES OF EMPIRICAL RATE AND R^2
TATTNALL COUNTY FULL DATA

READING

RAW SCORE

#147	RATE	
M	1.7	
F	-0.2	3.5
5%	-2.9	5.7
*	-14.2	18.7

#147	R^2	
M	33.6	
F	10.9	66.9
5%	0.4	92.7
*	0	100.0

GRADE EQUIVALENT

#152	RATE	
M	0.7	
F	0.4	1.1
5%	-0.4	1.8
*	-1.4	2.9

#152	R^2	
M	60.4	
F	24.5	84.9
5%	2.6	97.4
*	0	99.5

MATHEMATICS

RAW SCORE

#146	RATE	
M	3.0	
F	0.5	5.7
5%	-4.1	9.8
*	-11.0	12.6

#146	R^2	
M	25.1	
F	6.8	63.1
5%	1.0	95.2
*	0.0	98.0

GRADE EQUIVALENT

#152	RATE	
M	0.9	
F	0.6	1.4
5%	-0.3	2.0
*	-0.9	2.3

#152	R^2	
M	63.3	
F	37.7	88.6
5%	2.5	97.3
*	0.1	98.9

TABLE 4
STATISTICAL AND PSYCHOMETRIC QUANTITIES
TATTNALL COUNTY FULL DATA

READING

ESTIMATE	RAW SCORE	GRADE EQUIVALENT
t^o	10.796	15.417
κ	6.375	5.787
σ^2_θ	3.441	.089
$\sigma^2_{\xi(t^o)}$	139.882	2.978
$\rho(\theta)$	.281	.217
s.e.$(\hat{\theta})$	2.964	.567
γ	.766	.791
$\rho_{\theta W}$	-.085	.009

MATHEMATICS

	RAW SCORE*	GRADE EQUIVALENT
t^o	---	18.232
κ	---	.865
σ^2_θ	---	.077
$\sigma^2_{\xi(t^o)}$	---	.057
$\rho(\theta)$	---	.201
s.e.$(\hat{\theta})$	---	.552
γ	.769	.777
$\rho_{\theta W}$	---	-.223

Note: W is dichotomous gender variable.
* Warning: High SSRES causes computational error in some of these quantities.

the fourth lowest residual variance on reading. Student #109 has the highest learning rate in mathematics and the second lowest $\hat{\gamma}_p$ (which indicates that their relative position in the group is changing a lot over time).

Insert Tables 5 and 6 here

Tables 7 through 11 present similar analyses for the subset of 50 students that were used for analyses presented in Anderson v. Banks. In the following these 50 cases will be referred to as the "subset data." Table 7 shows that grade equivalents increase each year whereas raw scores increase until grade 12 and then drop. Table 8 shows, again, that GE's appear to work better for a straight-line growth model with these data. Tables 9, 10, and 11 repeat the analyses shown for the full data in Tables 4, 5 and 6, respectively. Gender was used for the background variable, W, in the computation of $\hat{\rho}_{\theta W}$.

3.1.3 Additional Growth Curve Analyses

During the completion of these analyses, several issues arose which gave rise to the subsequent analyses. First, it was recognized that it would be desirable to incorporate other more relevant background variables in the analyses if such variables could be identified. Second, it was realized that there existed considerable potential for topping out when using GE scores since the maximum raw score necessary to achieve the highest GE is relatively low (see Table 1). Related to this concern is the more general issue of which score metric is most useful for describing growth. Finally, it was beginning to become clear that the data were not consistent with the hypothesis that all

TABLE 5
TATTNALL COUNTY FULL DATA
CASES WITH EXTREME VALUES OF SELECTED QUANTITIES
RAW SCORES

READING $\hat{\theta}$	READING R^2	READING θ_{ε}^2	READING $\hat{\gamma}$		MATHEMATICS $\hat{\theta}$	MATHEMATICS R^2	MATHEMATICS θ_{ε}^2	MATHEMATICS $\hat{\gamma}$
034	012	057	017	LOW	196	019	171	152
196	144	042	034		152	161	028	109
149	108	074	093		167	184	086	097
131	128	109	149		052	038	062	052
171	089	032	172		183	090	041	158
052	074	190	195		089	094	036	168
048	093	075	193		149	028	047	178
063	021	002	188		075	109	119	192
093	109	196	192		097	123	196	012
017	057	017	187	HIGH	109	063	026	161

TABLE 6
TATTNALL COUNTY FULL DATA
CASES WITH EXTREME VALUES OF SELECTED QUANTITIES
GRADE EQUIVALENTS

READING $\hat{\theta}$	READING R^2	READING θ^2_ε	READING $\hat{\gamma}$		MATHEMATICS $\hat{\theta}$	MATHEMATICS R^2	MATHEMATICS θ^2_ε	MATHEMATICS $\hat{\gamma}$
196	061	045	093	LOW	196	144	081	040
149	189	064	149		152	147	111	097
131	011	052	048		161	128	017	108
192	183	030	196		150	158	019	152
187	139	040	131		192	019	054	089
018	054	087	185		063	044	188	003
023	064	138	195		089	111	117	004
101	030	011	159		116	041	029	028
048	045	075	193		097	014	125	001
093	052	002	003	HIGH	040	017	026	012

four testings had been done with the same test (as originally believed). Thus the data should perhaps not be taken "as is" for the analytical purpose of estimating growth. Subsequent analyses attempted to deal with these issues.

Insert Tables 7 through 11 here

3.1.3.1 Curricular Variables

First, an attempt was made to identify other suitable background variables for analysis. This was accomplished when (after considerable detective work) the tape file GAREC was located and investigated. As explained earlier, the variables ELEMREAD and FINEENG were selected as background variables for Reading Total and ELEMMATH and FINEMATH were chosen for use with Mathematics Total. Since the GE's had appeared most useful in the first analyses, the grade equivalent analyses were repeated using the GE values recorded in GAREC (these were the same as before except, as noted, that the grade 12 Reading Total GE's arbitrarily had been assigned a four to the right of the decimal rather than a zero).

For these analyses, attention was focused on the correlates of change. Systematic individual differences in growth were modeled by

$$E(\theta|W) = \mu_\theta + \beta_{\theta W}(W - \mu_W)$$

where W stands for an exogenous background variable. See Rogosa and Willett (1985) for a complete description of this model and its properties. Table 12 shows the estimated values of $\hat{\rho}_{\theta W}$ and several other quantities of interest. It will be noted that $\hat{\rho}_{\theta W}$ ranges from -.47 to -.63 . These values were contrary to

TABLE 7
MEANS AND STANDARD DEVIATIONS
TATTNALL COUNTY SUBSET DATA

READING

GRADE	RAW SCORE Mean (n = 47)	RAW SCORE S.D.	GRADE EQUIVALENT Mean (n = 50)	GRADE EQUIVALENT S.D.
9	34.4	12.7	7.3	2.7
10	34.6	14.4	7.7	2.8
11	37.1	13.3	8.4	2.8
12	36.3	12.1	9.4	2.2

MATHEMATICS

GRADE	RAW SCORE Mean (n = 49)	RAW SCORE S.D.	GRADE EQUIVALENT Mean (n = 50)	GRADE EQUIVALENT S.D.
9	36.9	18.3	7.2	2.6
10	44.5	20.2	8.5	3.0
11	52.3	19.0	9.6	2.8
12	44.8	15.2	10.0	2.2

TABLE 8
FIVE NUMBER SUMMARIES OF EMPIRICAL RATE AND R^2
TATTNALL COUNTY SUBSET DATA

READING

RAW SCORE

#47	RATE	
M	1.2	
F	-1.0	3.5
5%	-8.8	5.7
*	-14.2	6.1

#47	R^2	
M	40.0	
F	11.7	65.3
5%	0.4	95.3
*	0	100.0

GRADE EQUIVALENT

#50	RATE	
M	0.8	
F	0.2	1.1
5%	-0.5	1.9
*	-1.4	2.1

#50	R^2	
M	60.4	
F	29.5	81.7
5%	0.3	98.7
*	0	99.5

MATHEMATICS

RAW SCORE

#49	RATE	
M	3.2	
F	0.8	5.7
5%	-4.1	10.2
*	-11.0	12.6

#49	R^2	
M	3.2	
F	6.9	58.4
5%	1.1	94.8
*	0.02	97.5

GRADE EQUIVALENT

#50	RATE	
M	1.0	
F	0.7	1.3
5%	-0.2	2.0
*	-0.9	2.3

#50	R^2	
M	59.6	
F	38.1	84.4
5%	3.2	97.7
*	1.2	98.6

TABLE 9
STATISTICAL AND PSYCHOMETRIC QUANTITIES
TATTNALL COUNTY SUBSET DATA

READING

ESTIMATE	RAW SCORE	GRADE EQUIVALENT
t^o	11.002	12.337
κ	4.586	4.969
σ_θ^2	5.472	.180
$\sigma_\xi^2(t^o)$	115.096	4.439
$\rho(\hat{\theta})$	.372	.372
$s.e.(\hat{\theta})$	3.037	.551
γ	.742	.762
$\rho_{\theta W}$	.255	.237

MATHEMATICS

	RAW SCORE*	GRADE EQUIVALENT
t^o	---	22.615
κ	---	4.992
σ_θ^2	---	.031
$\sigma_\xi^2(t^o)$	---	.769
$\rho(\hat{\theta})$	---	.091
$s.e.(\hat{\theta})$	---	.555
γ	.788	.799
$\rho_{\theta W}$	---	.496

Note: W is dichotomous gender variable
* Warning: High SSRES causes computational error in some of these quantities.

TABLE 10
TATTNALL COUNTY SUBSET DATA
CASES WITH EXTREME VALUES OF SELECTED QUANTITIES

RAW SCORES

READING					MATHEMATICS			
$\hat{\theta}$	R^2	$\hat{\sigma}_\varepsilon^2$	$\hat{\gamma}$		$\hat{\theta}$	R^2	$\hat{\sigma}_\varepsilon^2$	$\hat{\gamma}$
034	012	057	034	LOW	196	019	041	109
196	089	042	048		052	057	025	089
131	027	109	106		096	189	109	052
061	138	032	023		131	138	001	010
025	037	040	013		043	043	009	096
106	052	194	155		180	009	061	159
163	163	087	159		010	014	045	012
180	106	138	163		040	025	036	054
052	109	034	180		089	041	196	023
048	057	196	194	HIGH	109	109	026	189

TABLE 11
TATTNALL COUNTY SUBSET DATA
CASES WITH EXTREME VALUES OF SELECTED QUANTITIES
GRADE EQUIVALENTS

READING					MATHEMATICS			
$\hat{\theta}$	R^2	$\hat{\sigma}_\varepsilon^2$	$\hat{\gamma}$		$\hat{\theta}$	R^2	$\hat{\sigma}_\varepsilon^2$	$\hat{\gamma}$
196	161	045	048	LOW	196	019	019	040
131	189	052	060		189	134	054	089
194	005	040	106		134	052	014	010
013	014	054	013		019	096	041	043
034	057	009	034		096	013	189	096
052	048	060	163		014	054	180	137
106	040	027	180		109	109	027	163
060	054	005	194		010	010	194	012
023	045	087	159		089	041	163	023
048	052	138	189	HIGH	040	014	026	060

expectation since they implied, for example, that more exposure to regular courses was associated with slower growth. It was suggested that perhaps these values were artifacts of the ceiling effect of the GE's. That is, students with high values of FINEENG and FINEMATH perhaps have higher values of GE to start with and thus can't exhibit much growth due to the low ceiling of the GE scores.

Insert Table 12 here

3.1.3.2 Analyses on Imputed Raw Scores.

Parallel to this investigation, we decided to try to reappraise, in the context of Tattnall County's data, the relative merits of raw scores and GE's for the purpose of studying academic growth. Also, as mentioned earlier, we needed to identify precisely which edition, form and level of the CAT had been used for each testing.

The plots from Section 1.3 show the form of the transformation of raw score into Grade Equivalemts for each grade and for each of Reading Total and Mathematics Total taken from the norms tables of the 1970 CAT/5A and the 1977 CAT/19C and CAT/19D. Figures 1-8 of this section show the "empirical" GE-raw score correspondence for the test scores of the individuals in the Tattnall County subset data. The first four plots show the reading scores and the next four the mathematics scores. After carefully comparing the normative plots with the empirical plots, checking the empirical data against both sets of norms tables and noting the percentage of agreements and disagreements, it was concluded that the first three testings

ERIC

TABLE 12

CORRELATES OF CHANGE
TATTNALL COUNTY SUBSET DATA
GRADE EQUIVALENTS

READING TOTAL WITH ELEMREAD OR FINEENG

ESTIMATE	ELEMREAD	FINEENG
$\rho_{\theta W}$	-.587	-.527
$\beta_{\theta W}$	-.344	-.026
$\rho_{\xi(t^o)W}$	.495	.568
$\beta_{\xi(t^o)W}$	1.378	.131

MATHEMATICS TOTAL WITH ELEMMATH OR FINEMATH

	ELEMREAD	FINEENG
$\rho_{\theta W}$	-.631	-.468
$\beta_{\theta W}$	-.216	-.014
$\rho_{\xi(t^o)W}$	.372	.213
$\beta_{\xi(t^o)W}$	.648	.032

most probably belonged to CAT 5A while the grade 12 testing appeared to be CAT/19D.

Insert Figures 1-8 here

<u>Imputation 1.</u> In order to avoid the truncated high end of the GE scale but still maintain some degree of comparability across the four years, it was decided to use the raw scores from the first three years and to impute a comparable raw score for the twelfth grade by replacing the CAT/19D raw score with the CAT/5A raw score which corresponded to the same GE. For the top GE of the scale, the raw score at the midpoint of the corresponding range of raw scores was imputed. Table 13 shows that when this is done, the large negative correlations with the curriculum indices disappear. That is, with grade equivalents correlations with the curricula indices (R,M) are (-.527,-.468) from Table 12, whereas these correlations become near zero for raw scores in Table 13.

Insert Table 13 here

<u>Imputation 2.</u> In an effort to further refine the raw scores, two other types of imputation were tried. The next type (Imputation 2) imputes raw scores from matching GE's exactly where possible just as in Imputation 1 described above. The difference lies in how we handled the top end of the GE scale. Imputation 2 used the norming table GE's and raw scores from the middle range of the table to extrapolate a raw score-GE

Figure 1

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF RGE9*RRS9   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

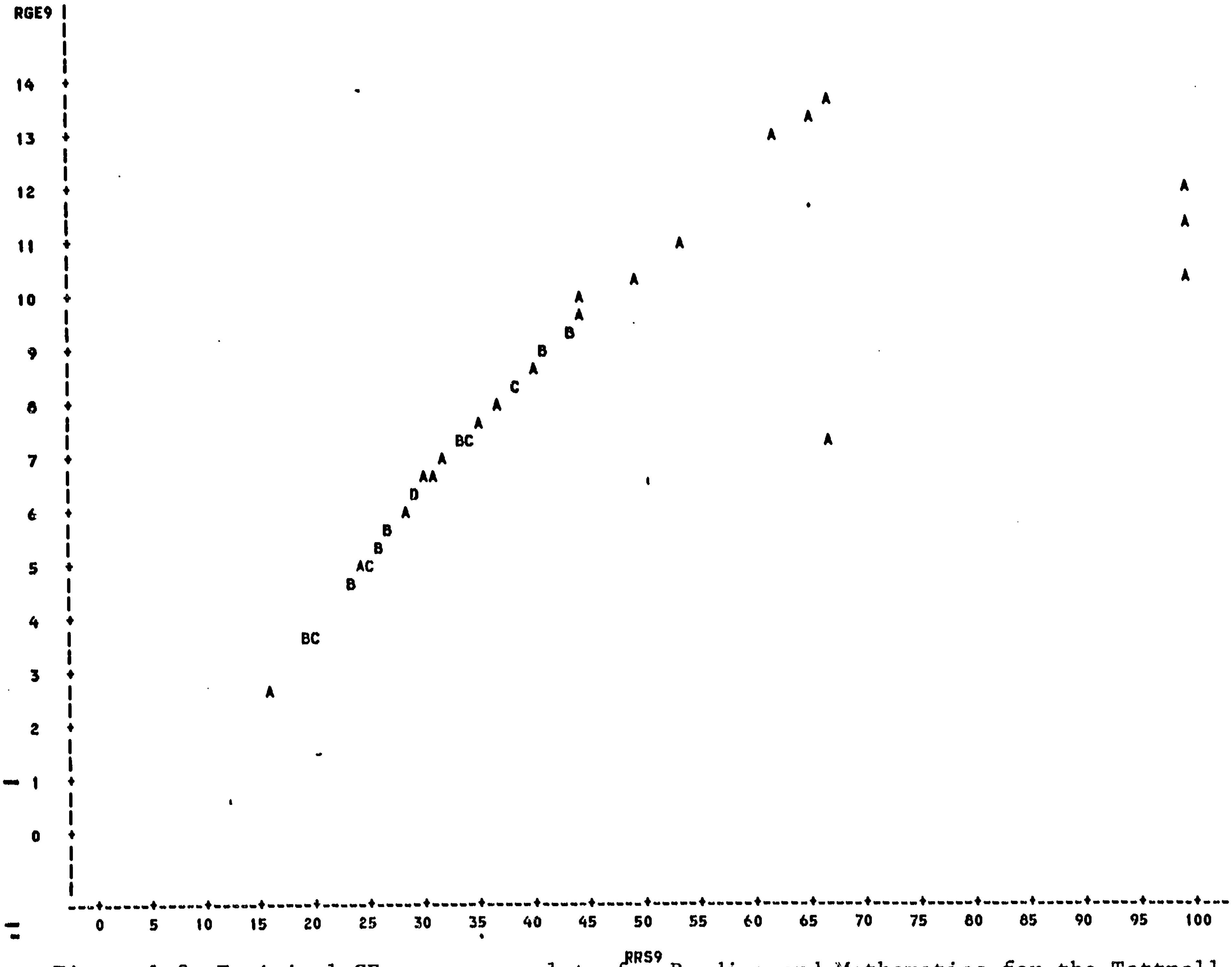


Figure 1-8. Empirical GE-raw score plots for Reading and Mathematics for the Tattnall County subset data.

Figure 2

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF RGE10*RRS10   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

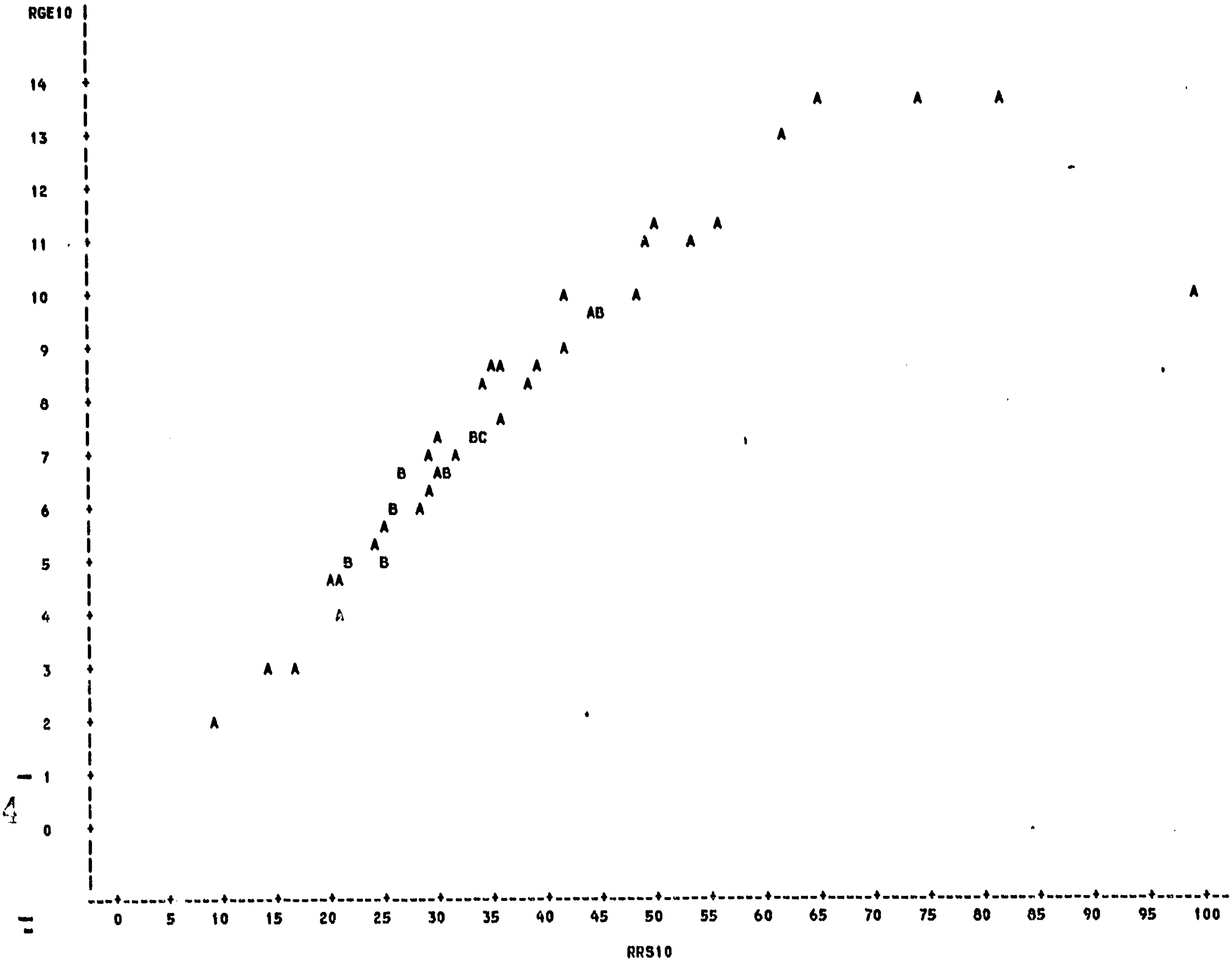


Figure 3

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF RGE11*RRS11 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

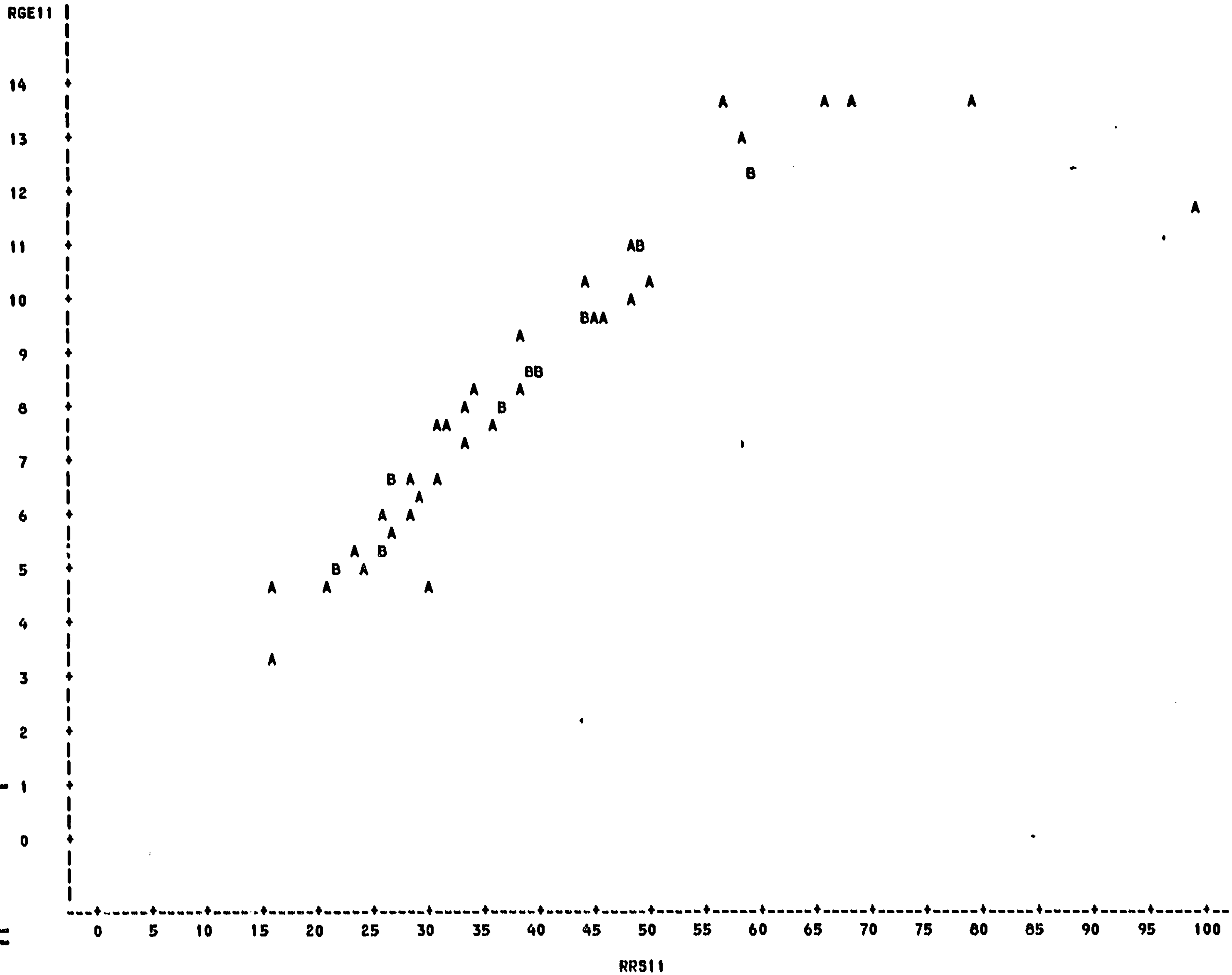
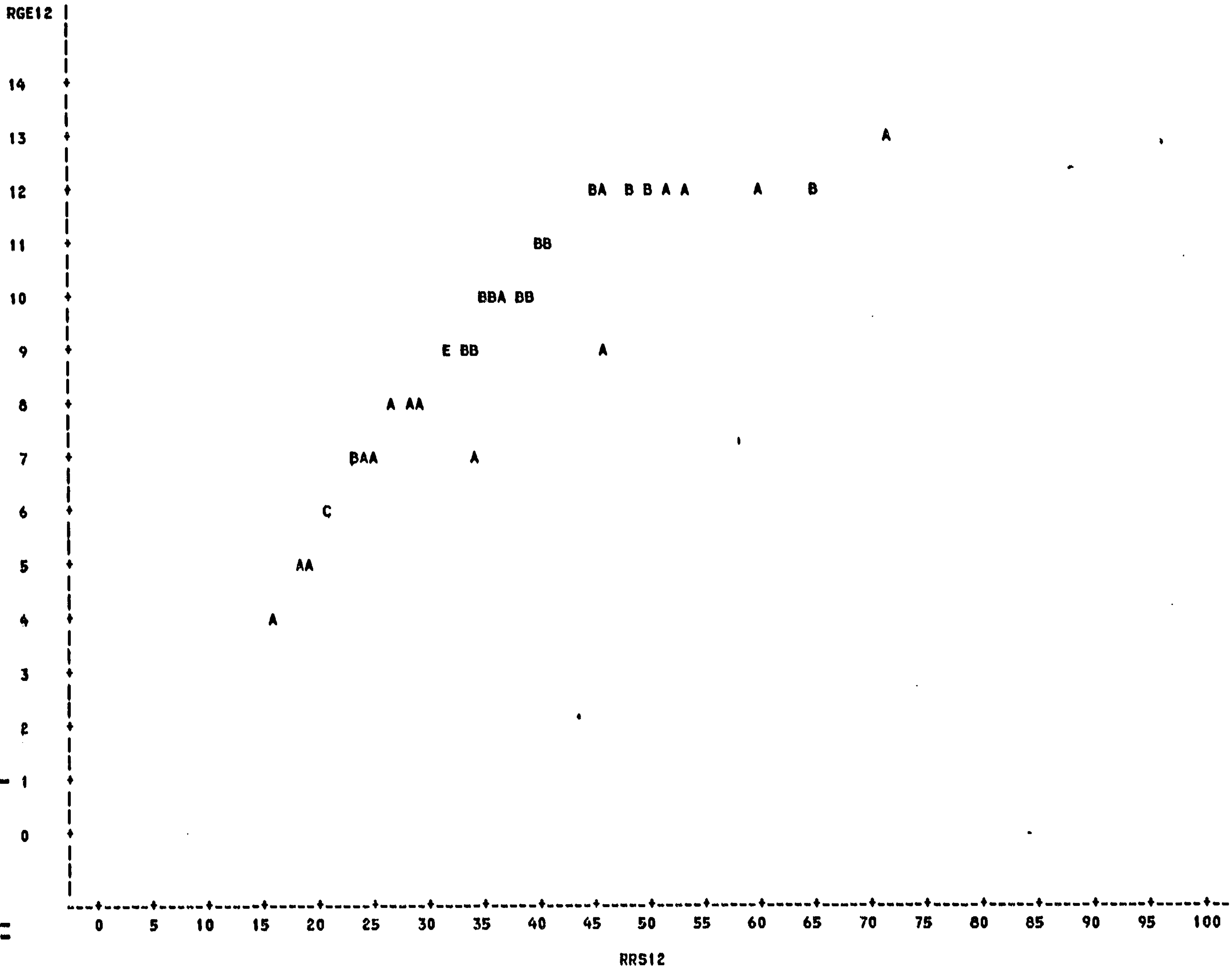


Figure 4

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF RGE12*RRS12 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



148

Figure 5

TATTNALL COUNTY SAMPLE DATA GE VS RS PLOTS

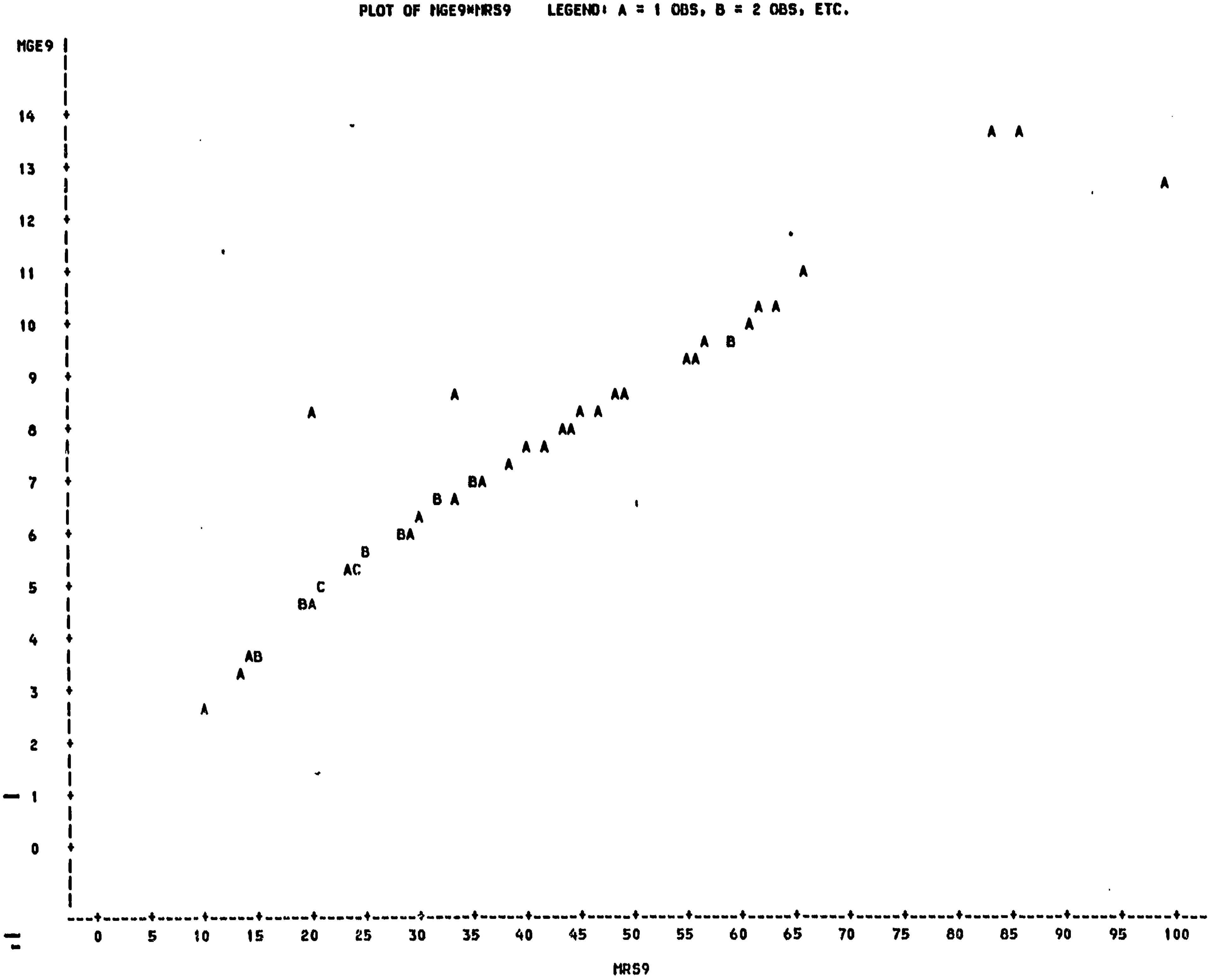


Figure 6

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

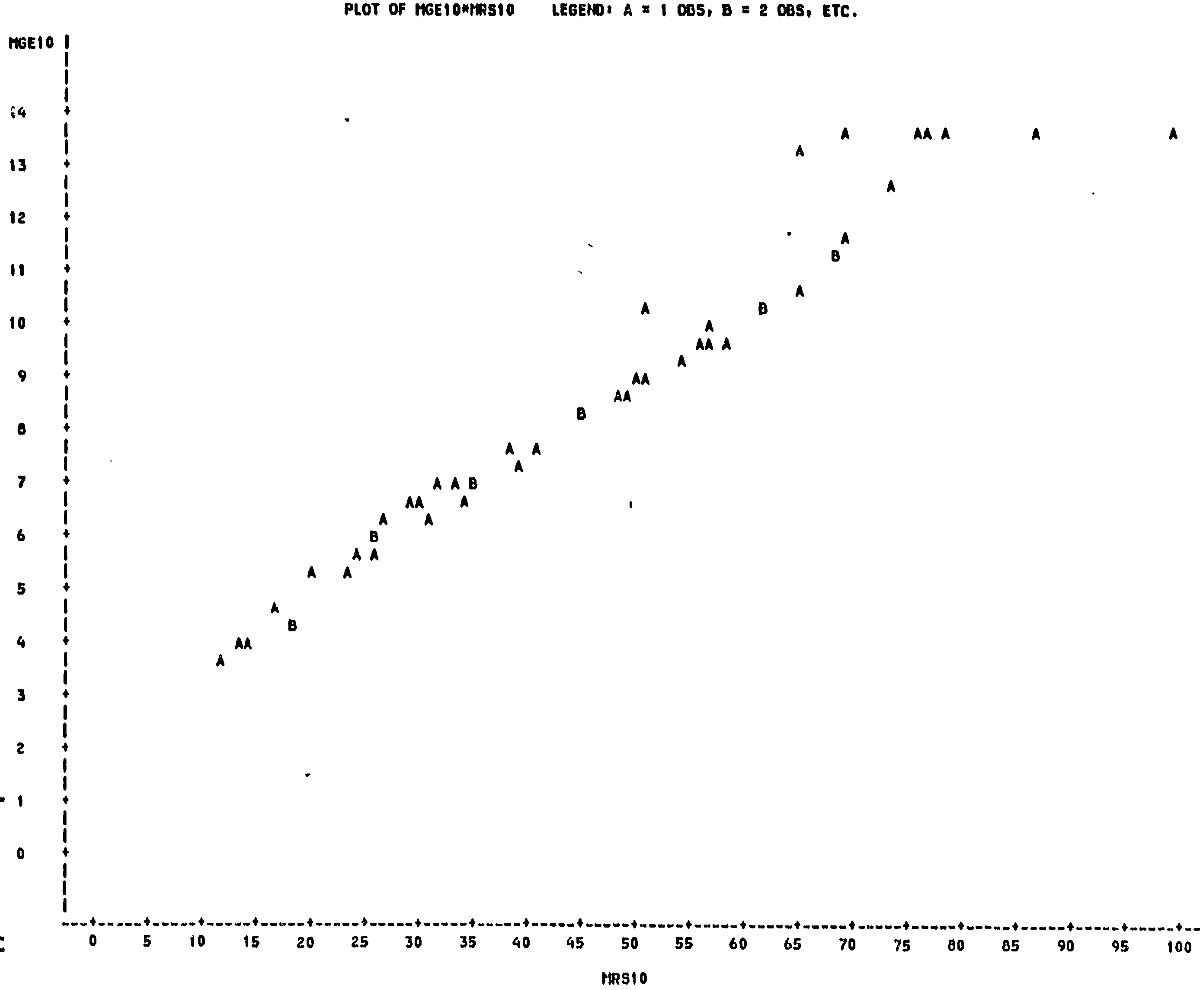


Figure 7

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF MGE11*MRS11 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

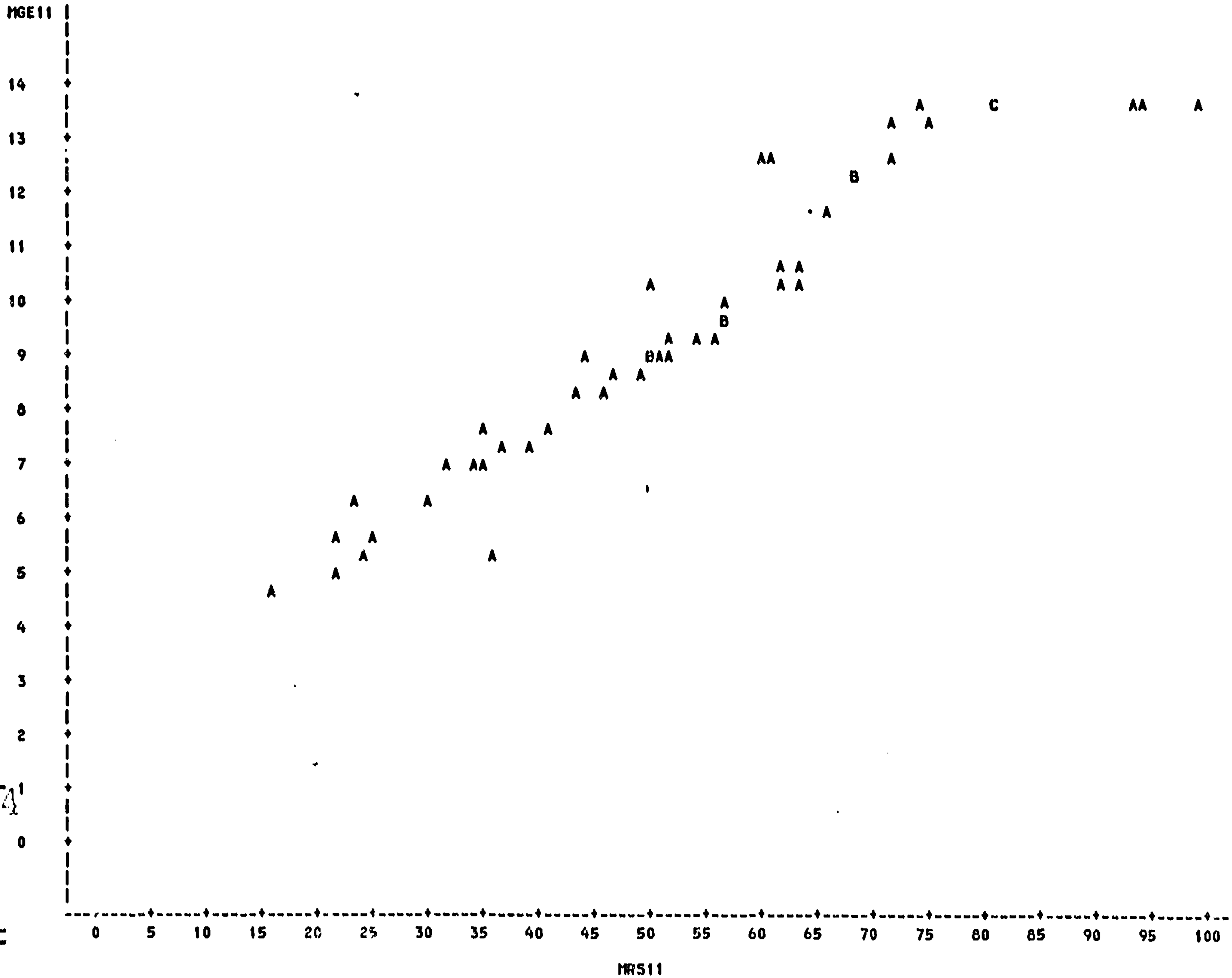


Figure 8

TATTNALL COUNTY SAMPLE DATA GE VS. RS PLOTS

PLOT OF MGE12*MRS12   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

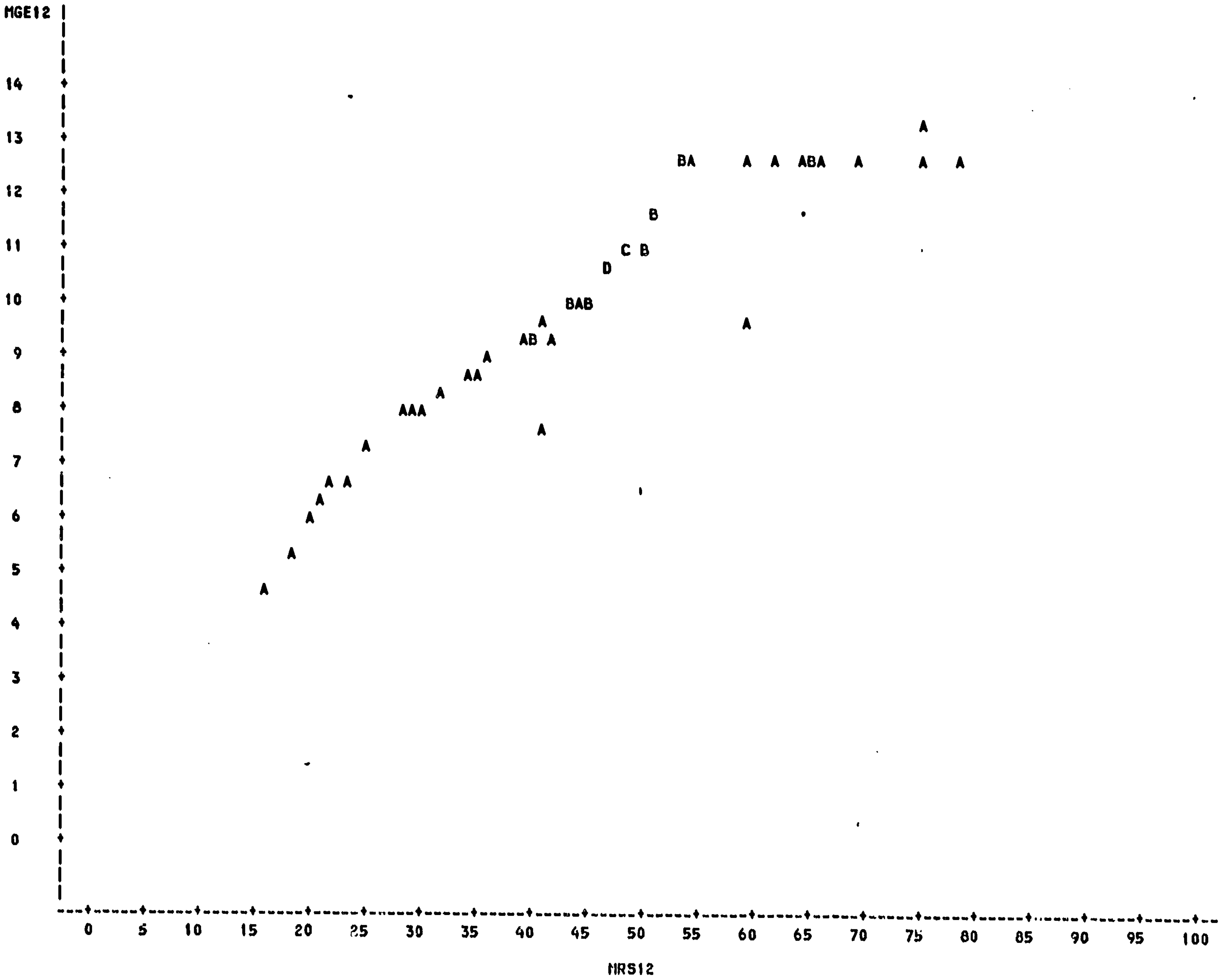


TABLE 13

CORRELATES OF CHANGE
TATTNALL COUNTY SUBSET DATA WITH IMPUTED RAW SCORES

READING

ESTIMATE	ELEMREAD	FINEENG
$\rho_{\theta W}$	-.412	-.097
$\beta_{\theta W}$	-1.017	-.020
$\rho_{\xi(t^o)W}$	.753	.754
$\beta_{\xi(t^o)W}$	11.733	.975

MATHEMATICS

	ELEMREAD	FINEENG
$\rho_{\theta W}$	-.398	-.032
$\beta_{\theta W}$	-1.323	-.009
$\rho_{\xi(t^o)W}$	.628	.478
$\beta_{\xi(t^o)W}$	11.552	.740

Note: Imputation 1 used

correspondence for the high end of the scale. Then twelfth grade raw scores were imputed by matching the predicted GE values of CAT/19D with those of CAT/5A.

Imputation 3. Imputation 3 was a different approach. Rather than use the GE's to achieve the imputation, we simply calculated percent-correct scores on CAT/5A and CAT/19D by dividing the scores on each by the respective total number of items. Then the raw score for CAT/19D was imputed as the CAT/5A raw score that corresponded to the same percent-correct score.

Table 14 compares the values of $\hat{\rho}_{\theta W}$ for the four sets of analyses (GE's and the three types of imputed raw scores). As can be seen, the negative correlations with FINEENG and FINEMATH disappear when raw scores are used, and the correlations with ELEMREAD and ELEMMATH decrease in magnitude except for ELEMMATH, raw score Imputation 3 where $\hat{\rho}_{\theta W}$ goes to -.705!

Insert Table 14 here

Table 15 compares the five number summaries of the squared multiple correlation for the individual growth curve fits (RSQ) for each of these four analyses. For Mathematics, raw score Imputations 1 and 2 clearly produce better fits than using the GE's while raw score Imputation 3 does not do as well. For Reading Total, it appears that the GE's conform slightly better to the straight-line growth model than do any of the raw score imputations.

Insert Table 15 here

3.1.4 Summary and conclusions

TABLE 14

COMPARISON OF ESTIMATED CORRELATES OF
RATE FROM GE AND RAW SCORE ANALYSES

READING		ELEMREAD	FINEENG
GE		-.587	-.527
RS:	Imputation 1	-.416	-.097
	Imputation 2	-.581	-.151
	Imputation 3	-.305	-.107

MATHEMATICS		ELEMMATH	FINEMATH
GE		-.631	-.468
RS:	Imputation 1	-.398	-.032
	Imputation 2	-.344	+.124
	Imputation 3	-.705	+.052

TABLE 15

FIVE NUMBER SUMMARIES OF R^2
FOR ANALYSIS OF TATTNALL COUNTY SUBSET DATA

		READING			MATHEMATICS	
GE	#43			#47		
	M	65.2		M	60	
	F	37.7	88.4	F	43.7	86.5
	5%	.4	98.5	5%	2.3	97.9
	*	0.	99.0	*	1.2	98.6
RS Imputation 1	#40			#42		
	M	59.5		M	75.0	
	F	33.0	78.9	F	49.8	93.3
	5%	1.9	97.9	5%	7.6	98.3
	*	.9	98.8	*	3.5	99.5
RS Imputation 2	#40			#42		
	M	59.5		M	75.0	
	F	33.0	85.0	F	49.8	92.2
	5%	1.9	97.9	5%	7.6	98.3
	*	.9	98.8	*	3.5	98.5
RS Imputation 3	#40			#42		
	M	53.7		M	55.7	
	F	22.4	80.1	F	29.5	77.9
	5%	1.5	98.4	5%	.9	99.5
	*	.2	99.5	*	.2	99.7

Analyses based on statistical estimation of growth curves were applied to the achievement data from the Tattnall County cases. As with many reanalyses perhaps more questions were raised than resolved. Clearly, the data and their analyses for the court case have myriad shortcomings; yet, these problems are even more profound in the other cases we examined. But none of the flaws in data and data analysis vitiate the central role of questions about student progress.

Quality of data. Although the data on student achievement and curricula are more extensive than in other cases, the serious flaws and ambiguities in these data render any inferences from their analyses open to debate. Even the court's statement about the actual tests that were administered (CAT 1970 and 1977 editions) are not totally consistent with the numerical values of raw scores and GE in the data we examined. Moreover, even unsophisticated examination of the data by the court revealed glaring discrepencies in the data:

> the test results for some of the students were plainly bizarre. For example defendent's Exhibit 52 shows CAT scores for 1978 Glennville seniors for the tenth eleventh and twelfth grade administrations of the test and the twelfth grade reexamination where necessary. Mickey A.'s reading score on the CAT was 3.8 [GE units]. One year later the score was 8.0. At the time of his senior retest, his score had regressed to 4.1. Renwick F. wen from a 4.3 in reading and 3.9 in math in the eleven grade to 10.0 and 9.8 less than two years later. Darlene J., according to Defendents' Exhibit 53 which gives the respective figures for Reidsville, jumped from 3.3 in reading and 5.8 in math in the tenth grade to scores of 9.9 and 9.2 in the twelfth. Kennie K.'s scores improved from 2.8 in reading and 4.6 in math in the tenth grade to 9.0 and 9.5 on the twelfth grade retest. Vernie A. at Reidsville obtained the following reading scores: 4.9, ninth grade; 2.5 tenth grade; and a whopping 11.5 in the eleventh grade. Progress was not steady [emphasis added].

> According to Defendents' Exhibit 12 showing the gains by year for 1980 Reidsville seniors in the math section, Talmadge F. showed a gain of 5.0 in his freshman year's CAT score. His score deteriorated each year thereafter. The corresponding exhibit for Glennville, Exhibit 14, shows Ricky B., after having regressed .1 in the ninth grade, showed an incredible gain of more than 7 grade levels in his tenth grade math score. (520 F.Supp. at 492).

In our own examination and analysis of the data we encountered additional discrepencies and confusions. Of additional note in the court's statement above is the examination of individual student histories and the emphasis on student progress (in GE metric).

Metric for achievement. The presentation to the court and the court's decision was entirely in terms of GE scores on the achievement tests. However our Table 14 shows that the choice of the metric for achievement (as discussed in section 1.3.1) can have a large effect on the conclusions about student progress that may be drawn from the data. In Table 14 the quantities of interest are estimates of the true correlation between student progress and remedial/regular curricula exposure, a key issue in the plaintiff's presentation of evidence. Thus the issue of a metric for achievement has considerable practical import because of possible effects on the results of the analysis.

Key questions about growth. Although there are considerable impediments to the analysis of student progress that arise in the reanalysis, questions about student progress remain highly pertinent.

For example:

1. How well or poorly are students progressing in the lower

tracks? This questions involves the descriptive analysis of student rates of progress.

2. Would students in the lower tracks have made better progress in regular rather than remedial courses? This question involves the relation between curriculum and student progress and also the effects of ability grouping.(cf. 540 F.Supp. 761 at 764).

3. Was the failure of many students to meet the graduation requirements (9.0 GE) largely a result of educational deficits built up before the ability grouping was implemented? I.e., was their failure unaffected by the district's policies? This question involves the correlation of change and initial status and possible adjustment for intitial differences in status.

LISTING OF SYSTEM FILE INFORMATION 06/21/85 PAGE 2

FILE GAREC (CREATION DATE = 09/04/81)

DOCUMENTATION FOR SPSS FILE 'GAREC '

LIST OF THE 1 SUBFILES COMPRISING THE FILE

GAREC N= 50

DOCUMENTATION FOR THE 66 VARIABLES IN THE FILE 'GAREC '

REL PC"	VARIABLE NAME	VARIABLE LABEL	MISSING VALUES	PRT FMT
1	SEQNUM		NONE	0
2	SUBFILE		NONE	A
3	CASWGT		NONE	4
4	IDNUM	STUDENT IDENTIFICATION NO	999.	0
5	ELEMREAD	GR 4-8 GPA	999.	0
6	ELEMLANG		999.	0
7	ELEMMATH	GR 4-8 GPA	999.	0
8	IOWALANG		999.	0
9	IOWAMATH		999.	0
10	GPAENG		999.	0
11	UENG		999.	0
12	CENG		999.	0
13	GPAMATH		999.	0
14	UMATH		999.	0
15	CMATH		999.	0
16	GPADREAD	DEV READING COURSES	999.	0
17	UDREAD		999.	0
18	CDREAD		999.	0

165 Addendum to Section 3.1. Listing of variables on file GAREC for Tattnall County subset data.

ERIC

DOCUMENTATION FOR THE 66 VARIABLES IN THE FILE 'GAREC '

REL POS	VARIABLE NAME	VARIABLE LABEL	MISSING VALUES	PRT FMT
19	GPADMATH	DEV MATH COURSES	999.	0
20	UDMATH		999.	0
21	CDMATH		999.	0
22	GPARREAD	REMEDIAL READING COURSES	999.	0
23	URREAD		999.	0
24	CRREAD		999.	0
25	GPARMATH	REMEDIAL MATH COURSES	999.	0
26	URMATH		999.	0
27	CRMATH		999.	0
28	TOTCRED		999.	0
29	MCTREAD	1. PASS 0. FAIL	999.	0
30	MCTMATH	1. PASS 0. FAIL	999.	0
31	SATVERB		999.	0
32	SATMATH		999.	0
33	RACE	0. WHITE 1. BLACK	999.	0
34	READ4	READING GRADE EQUIV QUARTILE MIDPOINT	999.	0
35	SEX	0. MALE 1. FEMALE	999.	0

Addendum, page 2.

DOCUMENTATION FOR THE 66 VARIABLES IN THE FILE 'GAREC '

REL POS	VARIABLE NAME	VARIABLE LABEL	MISSING VALUES	PRT FMT
36	SCHOOL		999.	0
37	MATH76	MATH GRADE EQUIVALENT AT 76	999.	0
38	READ76	READING GRADE EQUIVALENT AT 76	999.	0
39	MATH77		999.	0
40	READ77		999.	0
41	MATH78		999.	0
42	READ78		999.	0
43	MATH79		999.	0
44	READ79		999.	0
45	READPATT	ENGLISH COURSE PATTERN NO. 1-11	999.	0
46	MATHPATT	MATHEMATICS COURSE PATTERN NO. 1-19	999.	0
47	MATHMARK		999.	0
48	READMARK		999.	0
49	NORMMATH		999.	0
50	NORMREAD		999.	0
51	RMATH		999.	0
52	VARMATH		999.	0
53	RREAD		999.	0
54	VARREAD		999.	0
55	INTMATH		999.	0
56	SLOMATH		999.	0

Addendum, page 3.

ERIC

DOCUMENTATION FOR THE 66 VARIABLES IN THE FILE 'GAREC '

REL POS	VARIABLE NAME	VARIABLE LABEL	MISSING VALUES	PRT FMT
57	INTREAD		999.	0
58	SLOREAD		999.	0
59	RGREAD	RELATIVE GROWTH IN READING	999.	0
60	RGMATH	RELATIVE GROWTH IN MATH	999.	0
61	QREAD	ENTRY QUADRANT IN READING	999.	0
62	QMATH	ENTRY QUADRANT IN MATH	999.	0
63	GROSENG	GROSS CURRICULUM INDEX ENGLISH	999.	0
64	FINEENG	FINE CURRICULUM INDEX ENGLISH	999.	0
65	GROSMATH	GROSS CURRICULUM INDEX MATH	999.	0
66	FINEMATH	FINE CURRICULUM INDEX MATH	999.	0

Addendum, page 4.

3.2 CRITIQUE AND REANALYSIS: Scheelhaase v. Woodbury

This case concerns the non-renewal of an Iowa seventh grade English teacher's contract. The general issues of this case were described in Section 2.1, "Evaluation of Teacher Performance". Low test scores of the teacher's students on the ITBS and ITED were reasons the school board gave for her dismissal. The plaintiff prevailed in the district court: "...professional incompetence as indicated by low scholastic accomplishment of students on specified tests was arbitrary and capricious, since teacher's professional competency could not be determined solely on the basis of students' achievement on the tests, especially where the students maintained normal educational growth rates." (349 F.Supp. at 988-9) Although the decision was reversed in the Appeals Court to favor the School District, both courts seemed to accept the standard of normal educational progress as important evidence.

What exactly does "normal educational progress" mean in this instance? Scheelhaase's expert witness, Norman Ashby, implicitly defined this notion by the calculations he presented to the Court. The evidence submitted to the Court and his testimony are found in the Joint Appendix for Civil Case No 73-1067 (U.S. Court of Appeals, Eighth Circuit). This critique examines the data and analysis presented by Ashby and presents alternative statistical methods and their results.

3.2.1 Background and Available Data.

In addition to the published summary of the District and Appeals Courts decisions, we obtained documents from the United States District Court of Appeals for the Eighth Circuit in Iowa. These documents included the Joint Appendix (No. 73-1067) and the following briefs: Brief of the Iowa Association of School Boards as Amicus Curiae (filed 4/5/73); Brief of the NEA as Amicus Curiae (filed 6/4/73); Appellants' Brief and Argument (filed 4/9/73); and the Appellants' Reply Brief and Argument (filed 6/20/73). The Joint Appendix contains the available test data and summarizes Ashby's testimony.

All students Grades 3-8 in the Woodbury School District take the Iowa Test of Basic Skills (ITBS) annually. The test is administered each year in January, and scores are reported to the schools in percentile rank and grade equivalent (GE) scales.

Data Structure.

The data available to us were found in Exhibit II of the Joint Appendix. There are four subscales for the ITBS for which GEs were reported: Reading (R), Total Language (TL), Total Work Study Skills (TW), Total Arithmetic (TA), and also Total Composite Score (C). These scores were available in the Joint Appendix for an average over all students in each of the grades (3-8). A cohort is defined by the students' anticipated year of high school graduation. Data on four cohorts were available:

COHORT	YEAR IN SEVENTH GRADE
1972	1966-67
1973	1967-68
1974	1968-69
1975	1969-70

In evaluating the performance of an English teacher such as Scheelhaase, the Reading and Total Language scores are relevant. Thus the basic data can be represented for Reading scores as $R7X_i$ where cohort is represented by 7X = 72,73,74,75 and grade level by i=3,4,5,6,7,8. Thus, reading score for the 1972 cohort in grade 4 is denoted $R72_4$. The notation for total language is identical (with the substitution of TL for R). Recall that all scores were available as grade-equivalents (GE).

An average rate of growth for each cohort in each skill area (R,TL) was calculated as the difference in GE scores from third grade to eighth grade divided by 5: that is, for reading, the calculation was: $(R7X_8-R7X_3)/5 = GR7X$ for 7X = 72,73,74,75.

Then, within each skill area, each cohort's growth rates (e.g., GR7X) were averaged producing an overall growth average for each skill area, denoted for reading by GR. This overall average includes data for all the students in the school during the years 1966-1970, and was used as the index of normal educational growth.

To compare this index with student performance ascribable to Scheelhaase's instruction, Ashby calculated Scheelhaase's students' averages in two batches. Since the ITBS was administered in January, the 'Grade 7' administration reflects

Scheelhaase's influence for the six months immediately preceding, and the 'Grade 8' administration reflects her influence on the first half of the school year. Thus the two growth increments relevant to Scheelhaase's instruction were (for reading) $R7X_7 - R7X_6$ and $R7X_8 - R7X_7$. Each of these were then averaged over cohorts.

Comparisons were made between Scheelhaase's students' growth averages and the overall growth averages within her content area of reading and total language. Because the averages seemed comparable (i. e., they were not very disparate), Ashby asserted and the court agreed that Scheelhaase's students made normal educational progress. Table 1 summarizes the indices Ashby presented.

Insert Table 1 here

ERIC

TABLE 1

INDICES FROM SCHEELHAASE TESTIMONY

			SCHEELHAASE INCREMENTS			
COHORT	GR7X	GTL7X	R6-7	R7-8	TL6-7	TL7-8
72	.90	.86	.80	.90	.60	1.0
73	.96	.86	1.1	.70	.80	.70
74	.86	.82	.90	.90	.60	1.0
75	.92	.86	1.1	.80	.80	.90
AVE	.91	.85	.98	.83	.70	.90

Compare .91 vs 1/2 (.98 + .83)

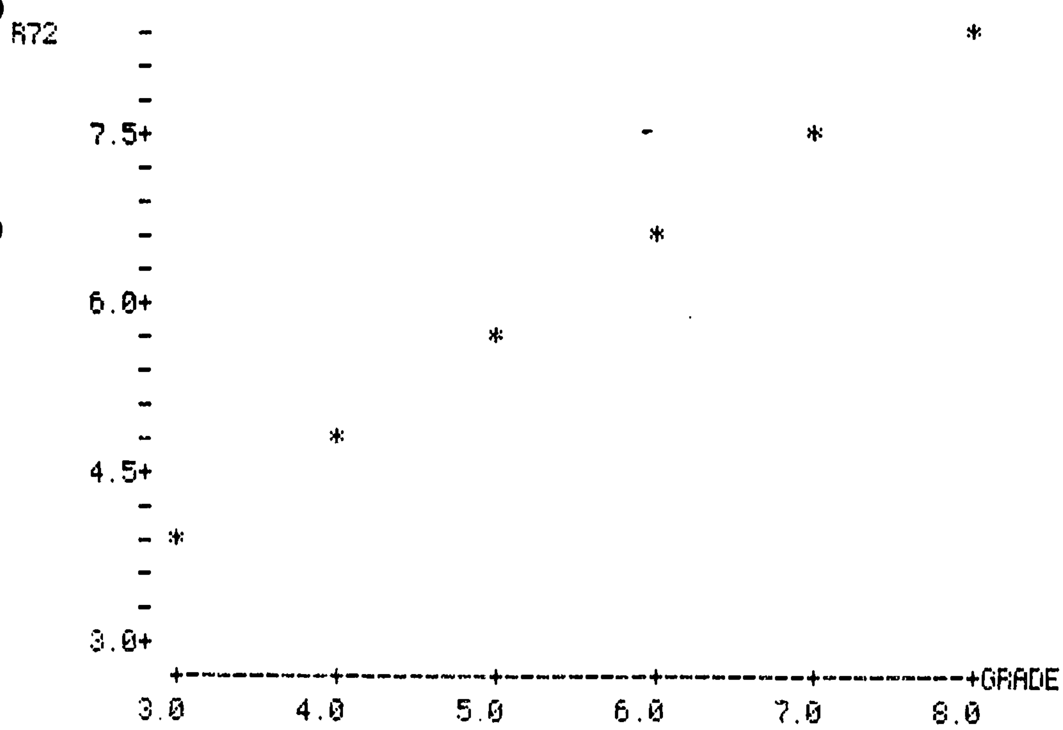
.85 vs 1/2 (.70 + .90)

3.2.2 Reanalysis of GE Data

The available data permit a longitudinal analysis of the progress of each cohort (72,73,74,75) in each of the relevant skill areas (i.e., R, TL) from third to eighth grade in the metric of grade equivalent scores (GE). The most appropriate initial examination of the data is to plot the GE scores against grade. Figure 1 contains these eight plots (4 cohorts and 2 subtests). Examination of the plots indicates that, at least for the average over students, the GE metric of constant rate of growth is valid. Fitting a straight-line growth curve to each of the plots yields results shown in Table 2.

Insert Figure 1 and Table 2 Here

Table 2 clearly shows a strong straight-line tendency in the GE scores. The squared multiple correlations for these fits all exceed .97. The slope of each straight-line fit represents the rate of growth for each cohort on each subtest over the span grades 3-8. This slope is an improvement as a growth index over the index used by Ashby in that the slope utilizes all the data (via the least-squares fit) instead of just utilizing the endpoints (grades 3 and 8). A comparison of Tables 2 and 1 does not show marked differences between the slope and Ashby's index, mainly because of the strong conformity of these data to a straight-line.



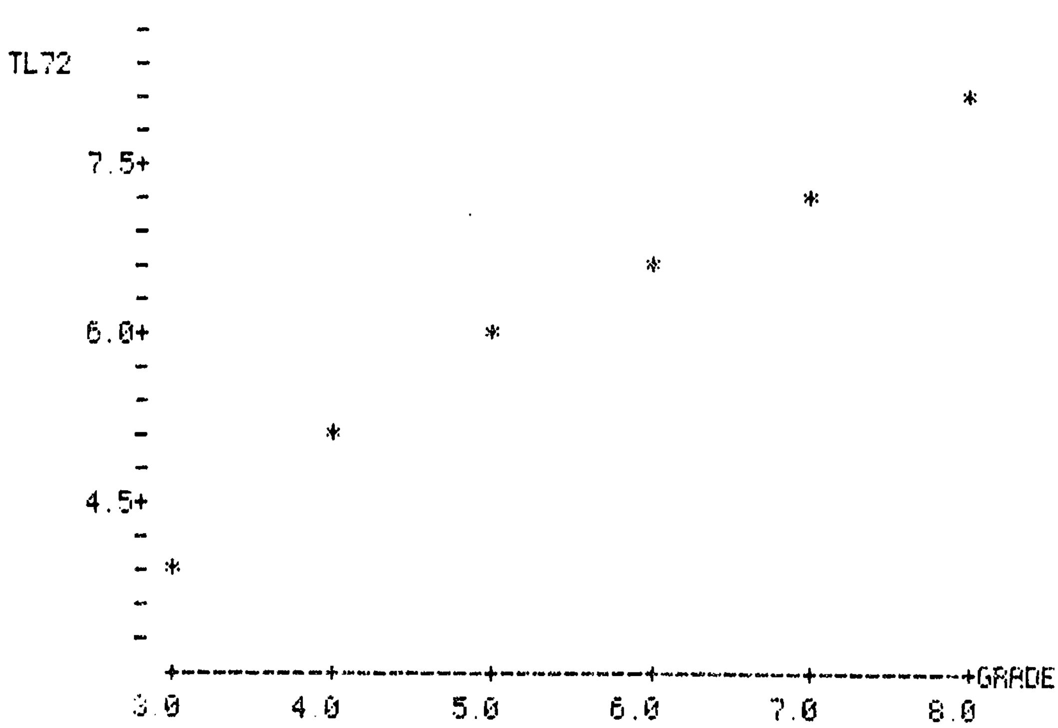
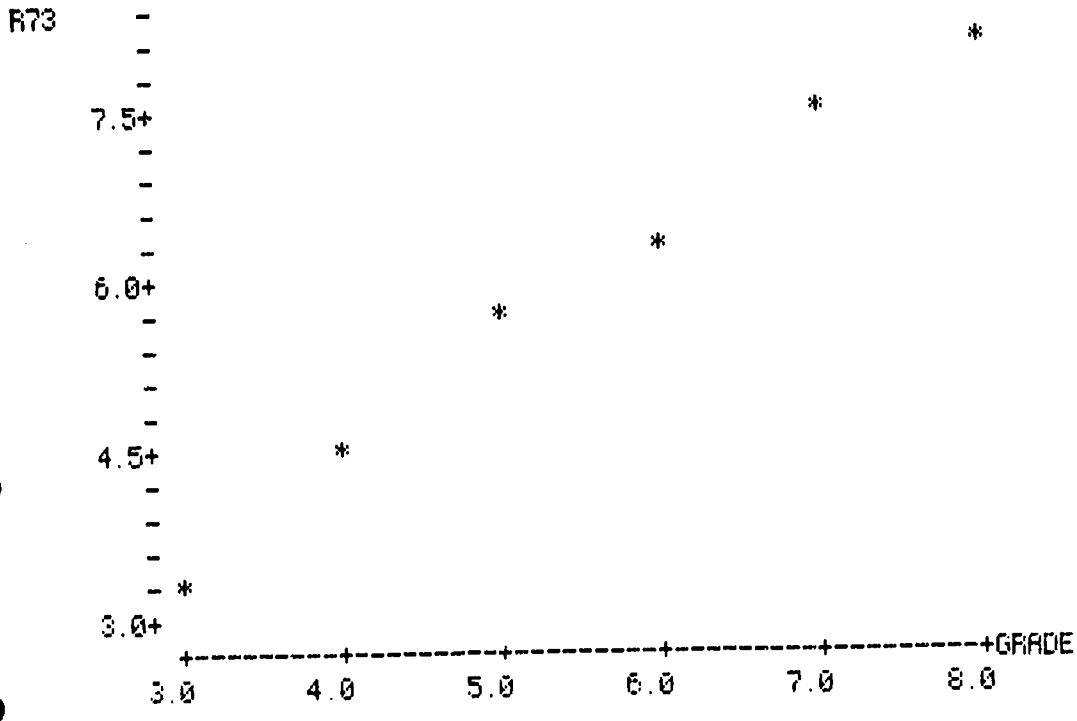


Figure 1. Plot of GE scores against grade level for Reading (R) and Total Language (TL) subtests for each of the four cohorts.



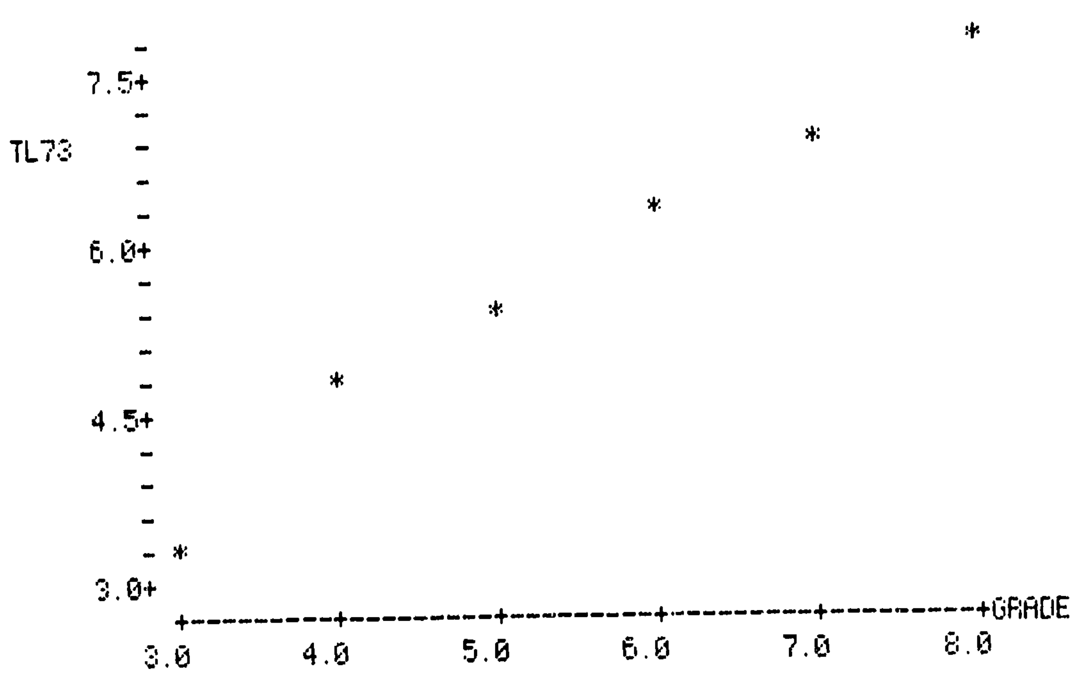
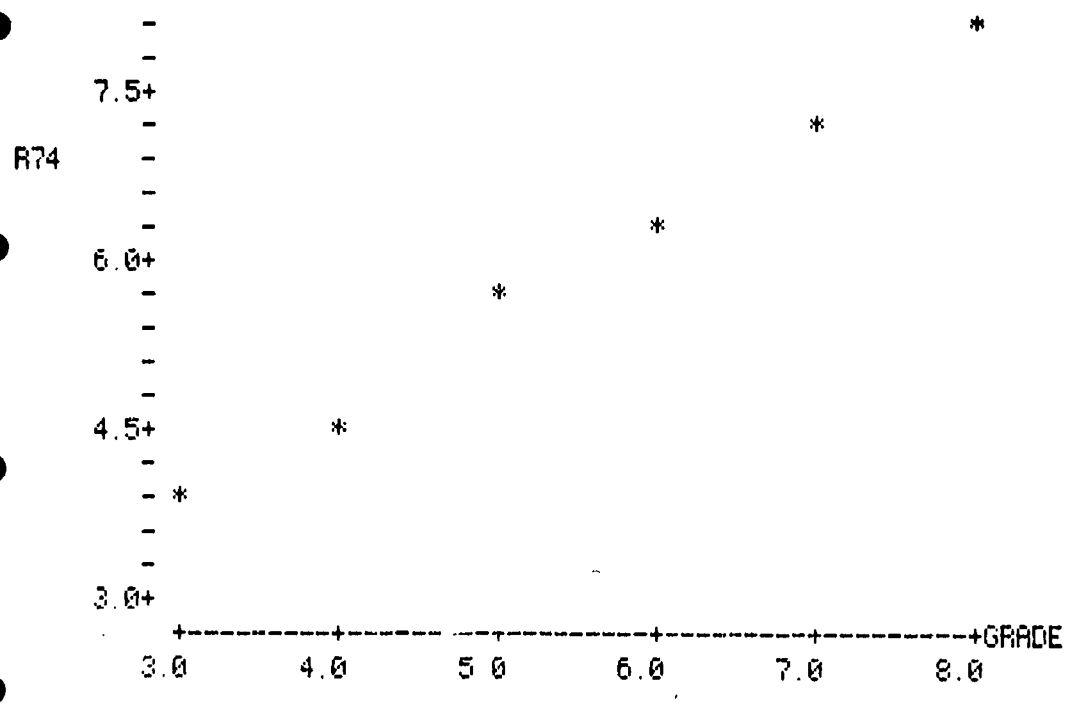


Figure 1, page 2.



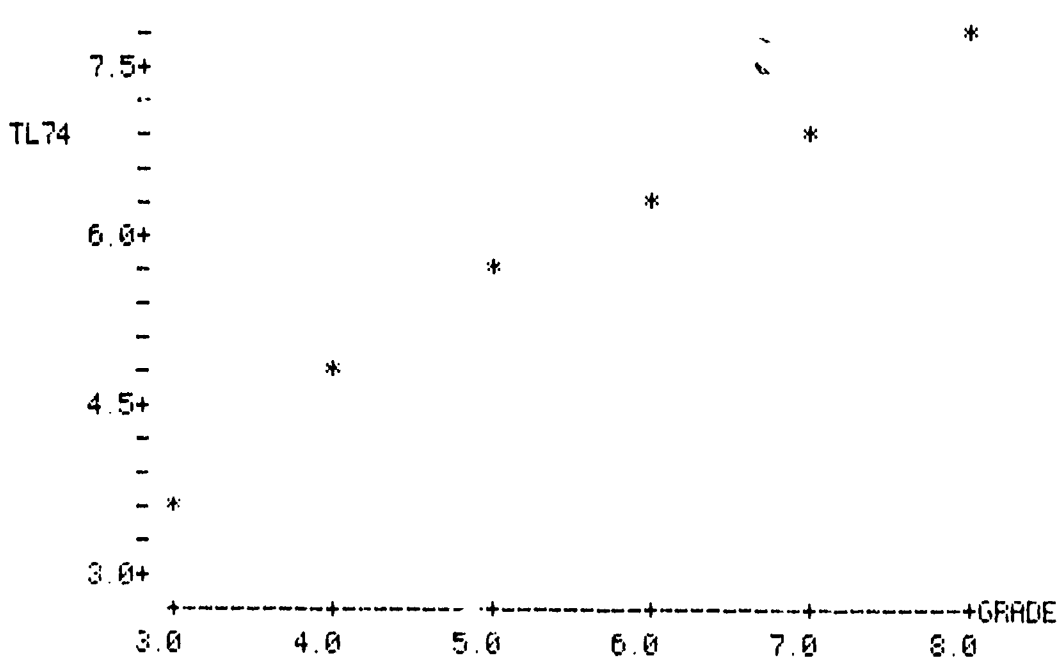
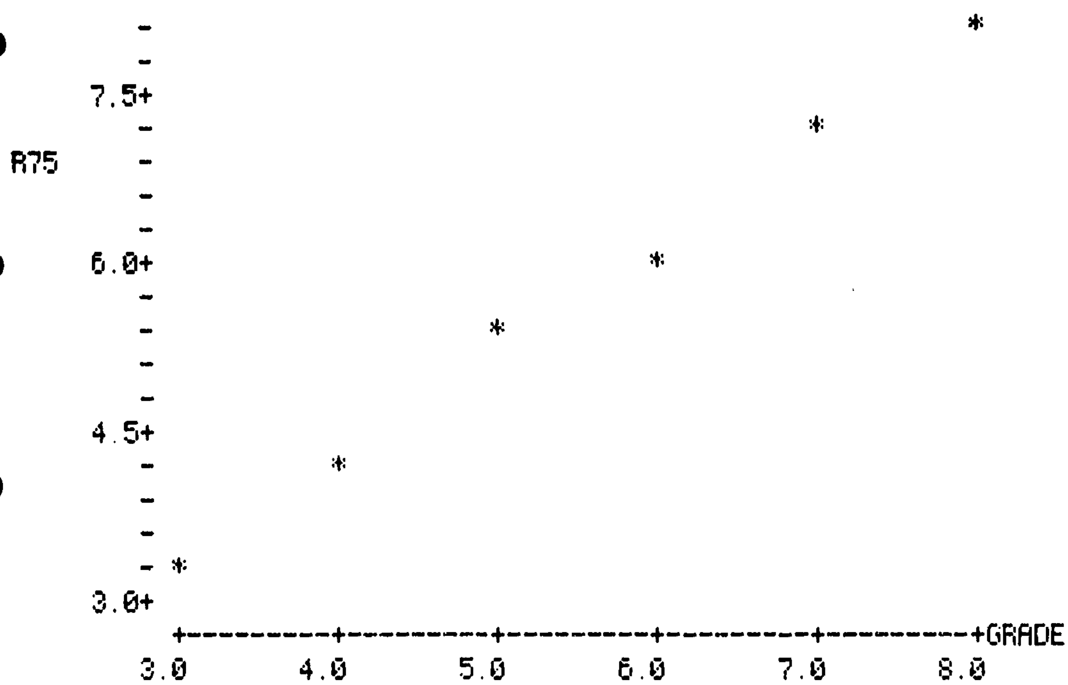


Figure 1, page 3.



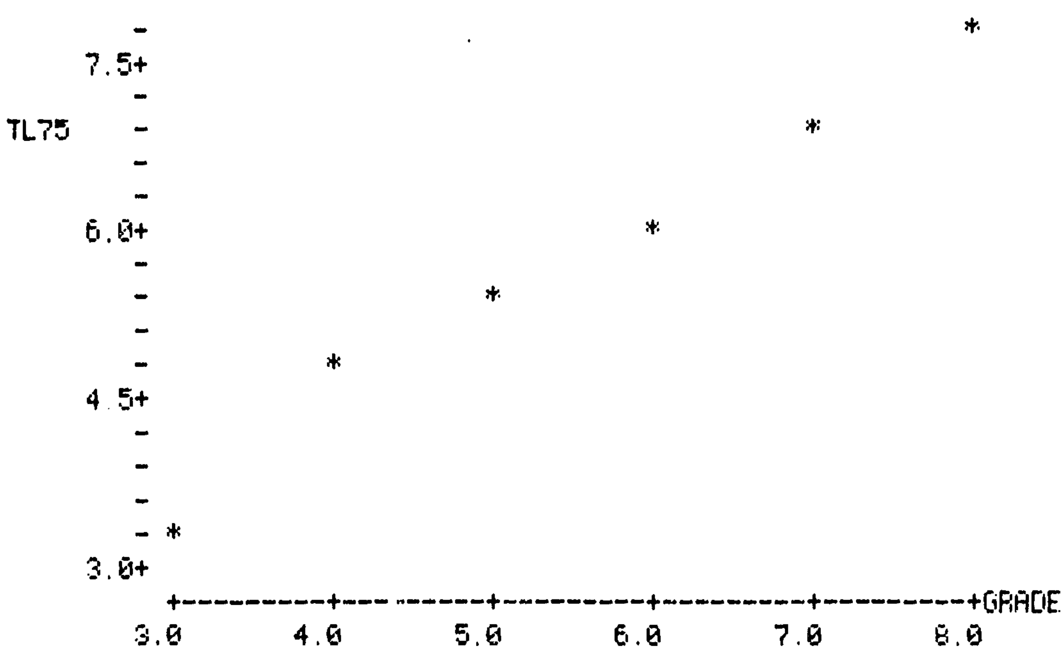


Figure 1, page 4.

TABLE 2

STRAIGHT-LINE FITS TO GRADE 3-8 DATA

	SLOPE	95% CI	RSQ
R72	.90	(.88, .92)	.999
R73	.97	(.90, 1.04)	.995
R74	.86	(.83, .89)	.999
R75	.93	(.88, .98)	.997
TL72	.82	(.74, .90)	.991
TL73	.82	(.70, .94)	.979
TL74	.77	(.67, .87)	.985
TL75	.81	(.72, .90)	.987

An important issue is a quantitative assessment of any discrepancy between a measure of "normal educational growth" and a measure of student progress attributable (in some manner) to Scheelhaase's instruction. In Ashby's testimony, such an assessment is handled highly informally: The comparison of Table 1 summary indices (.85 versus .80) is carried out with no statistical basis. The synopsis of the testimony reports that "roughly it (Scheelhaase' increment) appears to be fairly close to average" (p. 75 Joint Appendix).

The use of a regression estimate for the rate of improvement provides a natural way to assess a sizeable discrepancy. The 95% confidence interval for the slope in Table 2 defines an interval of plausible values for the rate of improvement. A natural comparison is whether the Scheelhaase relevant increments lie within the confidence interval for the slope.

Another problem with Ashby's growth index is that the data used to define "normal educational growth" included the data used to assess the student progress relevant to Scheelhaase's instruction. This overlap will certainly bias the analysis towards a conclusion that Scheelhaase's students do accomplish normal educational growth. An improved analysis would separate the data used to represent Scheelhaase's instruction. Such an analysis could use the four data points from grades 3-6 to establish a regression of GE on grade and then determine whether the GE scores from grades 7 or 8 fall within a 95% prediction interval based on this line. Table 3 reports these prediction intervals for each cohort and each subtest along with the observed GE scores. In each instance the GE score falls within its prediction interval. Note that some of the prediction intervals are rather wide. Thus this analysis cannot detect a deviation of the Grade 7 and 8 scores from the trend established in the lower grades.

Insert Table 3 Here

TABLE 3

PREDICTION INTERVALS FOR GRADE 7,8 DATA

	GRADE 7		GRADE 8	
	PREDICTION INTERVAL	OBSERVED DATA	PREDICTION INTERVAL	OBSERVED DATA
R72	(7.34,7.86)	7.5	(8.21,8.85)	8.4
R73	(6.70,8.31)	7.5	(7.54,9.50)	8.2
R74	(6.61.7.79)	7.2	(7.33,8.77)	8.1
R75	(6.10,8.10)	7.2	(6.82,9.24)	8.0
FL72	(6.64,8.56)	7.2	(7.33,9.67)	8.2
FL73	(5.01,9.39)	7.0	(5.43,10.75)	7.7
FL74	(5.74,8.66)	6.8	(6.26,9.82)	7.8
FL75	(5.23,8.77)	6.8	(5.70,10.00)	7.7

3.2.3 Examination of Alternative Functional Forms

Although grade-equivalent scores are constructed with the intent of yielding a constant rate-of-increase (i.e., 1 GE unit per grade), there are no guarantees that student scores will exhibit a constant-rate-of-change (i.e. straight-line growth). We used the Grade 3-8 data to examine the alternative functional forms for student progress. One common alternative to a constant-rate-of-change is a growth rate that depends on the student's current status--a constant relative rate of change conforms to exponential growth towards an asymptote. Table 4 presents both relative rate of change indices and the actual amount of change. The quantities in Table 4 are defined as follows (where X_i is the GE score for grade i):

$D_i = X_{i+1} - X_i$ Change between grade i+1 and i.

$S_i = D_i/X_i$ Relative rate of change.

$EX_i = \ln X_{i+1} - \ln X_i$ Rate of change in natural log units.

The latter two indices will be constant for exponential growth; the rate of change depends on status or distance to ceiling. Clearly, the data in Table 4 most closely conform to a constant-rate-of-change (straight-line growth).

Insert Table 4 here

TABLE 4

RATE OF CHANGE AND RELATIVE RATE OF CHANGE

		D_i		S_i		EX_i	
COHORT	GRADE	R	TL	R	TL	R	TL
1972	3	0.90	1.10	0.23	0.28	0.21	0.25
	4	0.90	0.90	0.19	0.18	0.17	0.17
	5	1.00	0.70	0.18	0.12	0.15	0.11
	6	0.80	0.60	0.12	0.09	0.11	0.08
	7	0.90	1.00	0.12	0.14	0.11	0.13
1973	3	1.00	1.5	0.29	0.44	0.26	0.36
	4	1.20	0.5	0.27	0.10	0.24	0.10
	5	0.80	0.8	0.14	0.15	0.13	0.14
	6	1.10	0.8	0.17	0.13	0.16	0.12
	7	0.70	0.7	0.09	0.10	0.09	0.10
1974	3	0.80	1.10	0.21	0.30	0.19	0.26
	4	1.00	0.90	0.22	0.19	0.20	0.17
	5	0.70	0.50	0.13	0.09	0.12	0.08
	6	0.90	0.60	0.14	0.10	0.13	0.09
	7	0.90	1.00	0.13	0.15	0.12	0.14
1975	3	0.80	1.30	0.24	0.38	0.21	0.32
	4	1.20	0.70	0.29	0.15	0.25	0.14
	5	0.70	0.60	0.13	0.11	0.12	0.11
	6	1.10	0.80	0.18	0.13	0.17	0.13
	7	0.80	0.90	0.11	0.13	0.11	0.12

Percentile Rank Data. The test results were also reported in "percentile rank" form. Percentile ranks are not in general a useful metric for assessing change over time. In addition these particular data were extremely volatile and their accuracy is unclear.

3.2.4. Shortcomings of the Data for Evaluating Teacher Performance.

1. The January testing date further obscures the attribution of student test performance to a particular teacher's instruction.

2. The lack of a comparison group (e.g. from similar districts) for progress in seventh - eighth grade instruction makes assessment of "normal educational progress" more difficult.

3. Group (school-averaged) data is less useful than individual data for assessing student progress.

4. Grade-equivalent metric may not be best (or even adequate) for analyzing student progress.

SECTION 4

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this section is to review some of the lessons learned from this research and to indicate possible future directions for methodological research. Section 4.1 restates some of the broad principles and procedures for the design and analysis of investigations of student progress. Section 4.2 sketches some unsolved methodological problems arising from some of the court cases examined.

4.1 IMPROVED DATA COLLECTION AND ANALYSIS

The previous sections of this report have demonstrated two conclusions: (1) questions about student progress play a key role in many court cases, and (2) current practice in measurement of achievement, data collection and management, and statistical analysis have serious deficiencies.

4.1.1 Data Collection and Management

School districts regularly assess students using group-administered achievement tests. Such testing represents a large investment in money and time for the schools, for administrators, for teachers and for students. Yet, local school agencies make relatively little use of the test data which they accumulate. In particular, test results are presented in a way that describes only the current status of students; the data are presented as a static "snapshot" of achievement without any link to prior levels of performance. Even the management of test data reflect these limitations. Whether the test results be stored as hard copy or electronically, the achievement data are typically organized as separate yearly files, which may be located on separate physical devices and even in separate geographical locations.

A key to the improved use of achievement test data is to use performance on repeated tests to describe student learning. A student's score at a single point in time cannot be used to measure learning; collecting together scores from previous testings is necessary for the analysis of student progress. A

ERIC

student's "cumulative folder" is organized in this manner, but these are rarely stored electronically nor uniformly maintained.

Design of data collection. The key design questions involve what data to collect and when (or how often). Thus (in addition to the typical question, How many individuals? which is often predetermined by the number of students and schools involved in the court case) design decisions include (i) the type and content of the achievement measure; (ii) the type of test score used to represent achievement; (iii) the number of and time between the assessments of achievement; and (iv) the type of exogenous measure(s) (e.g. curriculum, school program, initial status, demographic identification) to be linked (e.g., correlated) with individual or group progress. Standardized achievement tests were the predominant form of test instrument used in the court cases, in large part because these are the most avaliable measures. In many of the cases (e.g. Tattnall County) the court gave considerable attention and scrutiny to the appropriateness and relevance of these tests, particularly the match of the test and the content of the curriculum (curriculum validity following Debra P. v. Turlington). Clearly, an alternative to standardized achievement tests are tests locally produced, which can be tailored to the curricula and extended to content areas for which standardized instruments may not be available.

The type of achievement measures maintained in the student individual history is extremely consequential. Of the types of achievement scores reviewed in section 1.3.1 the scale score is the most appropriate measure with which to analyze student progress. The grade equivalent metric is the predominant type of score used, with advantages of simplicity of interpretation and implicit metric of expected progress. Raw scores also have some suitability for charting progress but only within the same level of the test. Particularly unsuitable for the analysis of student progress are percentile rank score and normal curve equivalents (NCE).

The number and spacing of the longitudinal observations is the most visible design decision (or constraint). Two observations, spaced a year apart constitute the most common design seen in the court cases. (Analyses, such as gain scores and regression methods, for the two-observation designs are descibed in Section 1.3.2.) As Rogosa et al. (1982, p. 744) state: "Two waves [observations] are better than one, but maybe not much better. Two data points provide meager information on individual change, and thus the measurement of change often will require more than the traditional pre-post data." Rogosa and Willett (1985, section 1.3) detail the shortcomings of two-wave data for the correlation of exogeneous variables (e.g., race or curriculum) with individual change. The statistical methods based on growth curve analysis require more than two observations, even for fitting the simplest, straight-line growth curve. The timing of the observations is also an important design descision. Measures collected a year apart predominate in the court cases because of traditional school testing practices. Denser observation schedules (e.g.

bimonthly) would make the collection of multiple (more than two) observations on each individual far more practical and relevant. Also the time of the school year at which the achievement measures are obtained can be important; for example, the annual January testing from Scheelhaase v. Woodbury (section 3.2) caused considerable difficulty in any assessment of a single teacher's performance (September-June).of teacher performance.

Often, a main purpose of the longitudinal analyses is to investigate a question about possible correlates of change. That is, the correlation between an exogenous variable and student rate of progress is of central interest. Whether race is related to rate of learning is a central question in the racial discrimination or unequal educational opportunity suits in section 2.2 (e.g., Serna v. Portales, section 2.2.2). The most common exogeneous variables in the court cases were curriculum or school program variables (In which curriculum or program to student make the most progress?), demographic or racial identification variables (Are minority student making the same progress as majority students?), or measures of initial status (Are the students with better initial standing making the most progress?). As seen in the Tattnall County reanalysis (section 3.2) the quality and choice of the exogenous variable is consequential.

4.1.2 Statistical Analysis

Statistical model for individual growth. Psychological learning theory and biological growth research provide a variety of complex models of individual growth, such as polynomial growth curves, logistic growth curves and simplex models. The simplest model, and the one used throughout this monograph, is the straight-line growth model,

$$\xi_p(t) = \xi_p(0) + \theta_p t \quad ,$$

where $\xi_p(t)$ is the true score of person p at time $t = 1, 2, \ldots, T$ and θ_p is the constant rate of change for person p. Thus, estimates of θ_p provide a simple index for individual rate of learning. The parameter, θ_p is closely related to the amount of true change; for example, in two-wave (or pre-post) data, true change defined as $\xi_p(t_2) - \xi_p(t_1)$ is equal to $\theta_p(t_2-t_1)$. In addition, the estimation of true scores $\xi_p(t)$ from observed scores $X_p(t)$ provides information about level as well as rate of learning.

The straight-line growth model is useful for heuristic reasons because of its simplicity, as it yields a simple index for individual rate of progress. In addition, Rogosa and Willett (1985) point out that, "in applications, straight-line growth serves as a useful approximation to actual growth processes" (p. 205). Moreover, when observations at only a few time-points are available, as with the Tattnall County data in section 3.1 where T = 4, the data may only justify the estimation of a constant rate of change. Although many uses of straight-line growth curves can be justified, non-linear growth functions may be crucially important in many applications, as in detecting effects of programs or teachers and in assessing deviations from "normal educational progress." Clearly, this is a worthwhile area for future statistical research and application; the tools and applications presented in this report serve at least as a useful first step.

Descriptive analyses of growth rates. The straight-line growth model does allow one to fruitfully compare individual learning for different individuals. For example, consider the four examples in Figure 1 of section 1.1. The figure shows the estimated individual learning "curves" obtained for four students by fitting a straight line to each student's individual longitudinal data obtained in the high school grades. Student (a) exhibits a rapid increase in reading achievement from a relatively low initial level of achievement. Student (b) starts out somewhat higher than student (a) and grows at a similar rate. Note that the observed scores are closely approximated by the fitted line. Student (c) exhibits slow growth from a relatively low initial achievement level, but the observed scores are more erratic than those for students (a) and (b). Finally, student (d) shows a high level of achievement, but no growth due to the ceiling effect of the test used.

When describing the learning of a group of individuals, the distribution, over individuals, of empirical rates of learning is informative. The five-number summary of empirical rates such as in Table 3 of section 3.1 is one useful way to describe both typical rates of learning and the degree of variability in rates of growth among individuals. Also of interest is the variability in θ_p, or estimates of θ_p. Thus σ^2_θ is a key quantity for investigation. Similarly, we may want to describe the variability in level of performance at each time, $\xi_p(t)$. As it turns out, $\sigma^2_{\xi(t)}$ has a functional dependence on time, and investigation of the form of this function leads naturally to the definition of a "centering" point and a scaling factor associated with the time scale. These have been denoted t^o and κ, respectively. (See Rogosa and Willett, 1985.) Both t^o and κ are properties of the particular collection of straight-line growth curves.

<u>Correlation of change and initial status</u>. Another quantity of central importance is the correlation between change, θ, and initial status, $\xi(t_I)$, where t_I indicates initial time of measurement. As discussed in Rogosa and Willett (1985), the choice of t_I is of critical importance because $\rho_{\xi(t)\theta}$ is functionally dependent on time. (The definitions of t^o and κ also arise naturally from an investigation of this dependence; see Rogosa and Willett, 1985.) Our statistical procedures provide a maximum likelihood estimate of the correlation between true rate of change and true initial status; the correlation between observed change and observed initial status is well-known to have a strong negative bias (see Rogosa et al. 1982). The correlation is used to investigate whether those with lowest initial status make the most progress (negative value) or those with the highest initial status make the most progress (positive value).

Correlation of exogenous variables with growth. More generally, there is interest in ways of describing systematic individual differences in growth, as indicated by the quantity $\rho_{\theta W}$ where W is some exogeneous background characteristic, for example, a characteristic of the school curriculum. The question addressed is whether students who experience certain values of W tend to exhibit more or less growth than students who experience other values of W. Our statistical procedures provide maximum likelihood estimates of this correlation, as in the correlations of progress with curricular indices in section 3.1.

In investigating systematic individual differences in growth, it is of course important to have a model for individual differences in growth. Rogosa and Willett (1985) state "Individual differences in growth exist when different individuals have different values of θ_p. Systematic individual differences in growth exist when individual differences in a growth parameter such as θ_p can be linked with one or more W's ." (p. 205) We use the simple representation

$$E(\theta|W) = \mu_\theta + \beta_{\theta W}(W - \mu_W) .$$

Thus non-zero values of $\beta_{\theta W}$ indicate that W is a predictor of growth. Alternatively, $\rho_{\theta W}$ is a useful summary quantity.

The typical procedure is to correlate the value of the background demographic variable or curricular variable with performance at a given time. That is, the cross sectional correlation is computed, sometimes for every occasion in time. For example, with a background variable, W , correlations of the test score with W at grade 9, 10, 11, 12 would be computed, and from these correlations conclusions about learning are attempted.

Rogosa and Willett (1985) have shown that such cross-sectional correlations cannot inform about student progress. To illustrate, consider a situation where the correlation between true rate of change and the background variable is zero. Then the correlation between the true test score, $\xi(t)$, and the demographic variable, W, at any one slice in time could be big or small. Consequently, $\rho_{\xi(t)W}$, really doesn't inform about systematic individual differences in learning. The reverse is true also. Consider a demographic variable for which $\rho_{\theta W}$ is large. Regardless, the correlation between the background variable and a test score at a specific time can be positive, zero, or negative depending upon the time chosen for the cross-sectional correlation. Obviously, no useful conclusions about learning can be drawn from the cross-sectional correlations.

<u>Consistency of individual differences.</u> The index γ was proposed by Foulkes and Davis (1981) as an index of tracking, and is defined as the probability that two randomly chosen growth curves do not intersect. High values of γ indicate high consistency of individual differences over time. Another way of interpreting γ is to note that high values of γ denote "the maintenance over time of relative ranking within the

response distribution" (Foulkes & Davis, 1981, p. 439). Thus γ indicates the stability of individual differences. If a collection of individual growth curves have a high value of γ, that indicates that individuals that started out relatively high maintain that advantage and individuals starting out low retain that disadvantage (regardless of the overall growth rate). Individuals with a low value of γ_p are those whose relative standing changes considerably over the time period.

ERIC

4.2 Further Methodological Problems

In the application of statistical analyses of student progress to the issues raised in the various court cases, a number of unsolved or unrecognized methodological problems became apparent. The discussion in this section may serve to direct some future technical research and development or help those involved in future court cases to recognize the limitations of empirical evidence or research capabilities.

4.2.1. Assessment of Normal Educational Progress

The determination of normal or adequate educational progress and the assessment of deviations from such are dominant methodological issues in many of the court cases reviewed in this project. Clearly, such determinations require substantial longitudinal data on student achievement and serious statistical analyses. In section 4.1 some general recommendations for the design and analysis of longitudinal data on student achievement are presented. In section 1.3.2, particularly section 1.3.2.3, previously used methods for determinations of adequate progress are reviewed, and the reanalyses in Section 3, particularly section 3.2, present and illustrate relevant statistical methods and problems.

Three basic questions for which neither answers or even complete methodological approaches are available are:

1. What is the functional form of normal educational progress? How does the functional form differ for

minority groups, different socio-economic levels, etc.?

2. In what test-score metric (raw score, GE, scale score) should expected status and progress in achievement be reported? Can these be linked to progress in curricula (e.g. basal reading sequence)?

3. What are useful methods for detecting and assessing deviations from normal educational progress (i.e., effects due to teachers, curricula, grouping practices?)

4.2.2. Classification Into Tracks by Standardized Test Scores

Ability grouping will necessarily rely upon imperfect classification schemes. Clearly, the more imperfect the classification mechanism the more difficult it will be for school districts to defend the "educational justification" for its ability grouping procedures. In particular, an important technical problem is to determine the error rates and optimal classification procedures when fallible measures such as standardized achievement tests (which contain errors of measurement) are used to form the ability groups. In the Dillon County case (section 2.4.1.3) the district was found to be in violation of Title VI of the Civil Rights Act as a result of its use of the total CTBS score for ability grouping. Particularly pursuasive to the judge was the testimony that bivariate classification--math scores used to form ability groups for math classes and reading scores used to form ability groups for reading classes-- would have greatly reduced the number of racially identifiable classrooms. Classification into groups on the basis of fallible scores has a small psychometric research literature which is summarized below. None of these procedures have been considered in the litigation involving ability grouping and the application of these procedures in school settings has not been explored to our knowledge.

Lord (1962) investigated cutting scores and errors of measurement for two selection variables (i.e. reading and math scores) assumed to be bivariate normal. Starting with the assumption that multiple cutting scores are optimum in the true score space, he determined the shape of the best selection region when fallible scores were used. Lord derived a formula for the cutting contour and illustrated the shape, direction and magnitude of the distortions produced by errors of measurement; his equation 19 expresses the form of the contours in terms of test-score statistics. Whereas the selection region defined by multiple cutting scores in the true score space is a quadrant cut of by a right angle with sides parallel to the co-ordinate axes, the best selection region based on the fallible observed scores is a curve resembling a hyperbola (Lord, Figure 1). Lord concludes that "other things being equal, the lower the reliability of the predictors, or the higher the correlation between them, the greater the difference between the multiple-cutting scores selection region and the optimum selection region" (p. 29).

More recently, Huynh (1982) presented a Bayesian approach to the determination of cutting contours when multiple test scores are used in the specific classification context of granting or denying mastery. He also discussed the influence of the loss ratio on the cutting contour and the distortions due to errors of measurement. Huynh illustrates his methods by the constuction of cutting contours for bivariate scores. Huynh describes the effects of errors of measurement by plotting cutting contours for bivariate normal X_1 , X_2 having different values of a common reliability coefficient, ρ (Huynh, Figure 2). In earlier papers Huynh (1976, 1977) considers a Beta-Binomial model for classifications of examinees into two groups (mastery, non-mastery).

4.2.3. Assessment of (Differential) Mobility Among Tracks

In racial discrimination or equal educational opportunity cases involving ability grouping or other classifications of students, issues of mobility among the tracks, especially differential mobility for majority and minority students, are central. Section 1.3.2.5 explores some issues and methods in improvement in classification status. Section 2.4.1 provides examples of cases involving ability grouping; differential mobility is illustrated in Figure 1 of section 2.4. Section 2.4.2 discusses special education cases. However, no criteria and little methodology exist for the systematic study of student mobility among classifications. For example, the Office of Civil Rights standards for racial identifiability of classrooms (± 20 percent) has no clear application in this context. Key questions include:

1. How can standards be set for mobility among tracks?

2. How can differential mobility among groups (e.g. minority and majority students) be quantified?

3. How can measures of progress in achievement be used to indicate reclassification of students (e.g., remedial to regular classrooms or special education to less restictive environment)?

REFERENCES

1. Cited Research Literature

Findley, W. (1979). Civil rights and elementary/secondary school measurement. New Directions for Testing and Measurement, 3, 79-85.

Findley, W., & Bryan, M. (1975). The pros and cons of ability grouping. Phi Delta Kappan, Fastback #66. Bloomingdale, IN: The Phi Delta Kappa Educational Foundation.

Haertel, E. (1984). Evaluating teachers by evaluating students: Problems and solutions in implementing some mandates of California's SB 813. Paper presented at the joint meeting of the Evaluation Network and the Evaluation Research Society, October 12, 1984, San Francisco.

Huynh, H. (1976). Statistical consideration of mastery scores. Psychometrika, 41(1), 65-78.

Huynh, H. (1977). Two simple classes of mastery scores based on the Beta-Binomial model. Psychometrika, 42(4), 601-608.

Huynh, H. (1982). A Bayesian procedure for mastery decisions based on multivariate normal test data. Psychometrika, 47(4), 309-319.

Lord, F.M. (1962). Cutting scores and errors of measurement. Psychometrika, 27(1), 19-30.

Mills, R., and Bryan, M. (1976). Testing ... grouping: The new segregation in Southern schools? Atlanta: Southern Regional Council.

Novick, M. (1982). Ability testing: Federal guidelines and preferential standards. In Ability testing: Uses, consequences, and controversies, Part II. Widger and Garner (eds.). Washington: National Academy Press.

Oakes, J. (1985). Keeping track: How schools structure inequality. New Haven: Yale University Press.

Rogosa, D.R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychologial Bulletin, 92(3), 726-748.

Rogosa, D.R. and Willett, J.B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50(2), 203-228.

Tallmadge, G.K., and Horst, D.P. (1976). A practical guide to measuring project impact on student achievement. Washington, D.C.: U.S. Government Printing Office.

Tallmadge, G.K., and Wood, C.W. (1976). User's Guide: ESEA Title 1 Evaluation and Reporting System. Mountain View, CA: RMC Research Corporation.

Test Coordinator's Handbook of the CAT, Levels 10-19 (1978). Monterey, California: CTB/McGraw-Hill.

White, O.R. and Haring, N.G. (1980). Exceptional teaching. Second Edition. Columbus, Ohio: Charles E. Merrill Publishing Co.

Wilson, B.J. and Schmits, D.W. (1978). What's new in ability grouping? Phi Delta Kappa, 59(8), 535-536.

2. Cited Court Cases and Related Legal Materials

Anderson v. Banks and Johnson v. Sikes (aka Tattnall County). U.S. District Court, Georgia. 520 F.Supp. 472 (1981); 540 F.Supp. 761 (1982); Court of Appeals. 730 F.2d 644 (1984).

Berry v. School District of Benton Harbor. U.S. District Court, Michigan. 515 F.Supp. 344 (1981).

California Senate Bill 813: The Hughes-Hart Educational Reform Act of 1983.

Carlin et al. v. Board of Education of San Diego Unified School District. No. 303800, Order Re integration Plan, 1980-81.

San Diego City Schools Planning, Research, and Evaluation Division, Report Number 425. "Testing Results from Minority Isolated Schools. Spring, 1986". September 2, 1986.

Debra P. v. Turlington. 474 F.Supp. 244 (1979); 644 F.2d 397 (1981); 564 F.Supp. 177 (1983).

Dillon County. Administrative Proceeding in the United States Department of Education, Initial Decision. Docket number 84-VI-16.

Georgia State Conference of NAACP v. State of Georgia. U.S. District Court, Savannah. 99 F.R.D. 16 (1983).

Marshall v. State of Georgia. Civil Case Judgement. CV482-233.

Hobson v. Hansen. U.S. District Court, Washington, D.C. 269 F.Supp. 401 (1967); 327 F.Supp. 844 (1971).

Hoffman v. Board of Education of City of New York. Appellate Division. 410 N.Y.S.2d 99 (1978).

Johnson v. San Francico Unified School District. 339 F.Supp. 1315 (1971).

Public Law 94-142. Education for All Handicapped Children Act of 1975. 89 Stat. 773.

San Francisco NAACP v. San Francisco Unified School District. U.S. District Court, California. 576 F.Supp. 34 (1983).

Consent Decree. San Francisco NAACP v. S.F.U.S.D. Civil No. C-78-1445 WHO. U.S. District Court for the Northern District of California.

Stipulation pursuant to Consent Decree Paragraph 50 and order modifying consent Decree. San Francisco NAACP v. S.F.U.S.D. Civil No. C-78-1445 WHO. U.S. District Court for the Northern District of California.

State Defendants' 1983-84 Annual Report pursuant to Paragraph 44 of the Consent Decree. San Francisco NAACP v. S.F.U.S.D. Civil No. C-78-1445 WHO.

State Defendants' 1985-86 Annual Report pursuant to Paragraph 44 of the Consent Decree. San Francisco NAACP v. S.F.U.S.D. Civil No. C-78-1445 WHO.

Scheelhaase v. Woodbury. District Court, Iowa. 349 F.Supp. 988 (1972); Court of Appeals 488 F.2d 237 (1974).

Brief of the NEA as Amicus Curiae. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Brief of the Iowa Association of School Boards as Amicus Curiae. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Joint Appendix. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Petition for Rehearing. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Appellants' Brief and Argument. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Appellee's Brief and Argument. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Appellants' Reply Brief and Argument. Scheelhaase v. Woodbury. No. 73-1067. U.S. Court of Appeals for the Eighth District.

Serna v. Portales. District Court, New Mexico. 351 F.Supp. 1279 (1972); Court of Appeals 499 F.2d 1147 (1974).

3. Additional Relevant References

Baldus, D.C. and Cole, J.W. (1980). Statistical proof of discrimination. New York: McGraw Hill.

Cohn, R.M. (1980). On the use of statistics in employment discrimination cases. Indiana Law Journal, 55, 493-513.

Cohn, R.M. (1980). Statistical laws and the use of statistics in law: A rejoinder to Professor Shoben. Indiana Law Journal, 55, 537-549.

Downton, F. (1982). Legal probability and statistics. Journal of the Royal Statistical Society, 145(4), 395-402.

Fienberg, S.E., and Straf, M.L. (1982). Statistical assessments as evidence. Journal of the Royal Statistical Society, 145(4), 410-421.

Finkelstein, M. (1978). Quantitative Methods in Law. Studies in the Application of Mathematical Probability and Statistics to Legal Problems. New York: The Free Press.

Finkelstein, M. (1980). The judicial reception of multiple regression studies in race and sex discrimination cases. Columbia Law Review, 80, 737-754.

Fisher, F.M. (1980). Multiple regression in legal proceedings. Columbia Law Review, 80, 702-736.

Gable, R.K. & Iwanicki, E.F. (1986). The longitudinal effects of a voluntary school desegregation program on the basic skill progress of participants. Metropolitan Education, 1, 65-78.

Gardner, E. (1982). Some aspects of the use and misuse of standarized aptitude and achievement tests. In Ability testing: Uses, consequences, and controversies, Part II. Widger and Garner (eds.). Washington: National Academy Press.

Hollander, P. (1982). Legal context of educational testing. In Ability testing: Uses, consequences, and controversies, Part II. Widger and Garner (eds.). Washington: National Academy Press.

Kaye, D. (1980). Naked statistical evidence. Yale Law Journal, 89, 601-611.

Kaye, D. (1982). Statistical evidence of discrimination. Journal of the American Statistical Association, 77, 773-792.

Novick, M. (1977). The influence of the law on professional measurement standards. Educational Measurement and the Law: Proceedings of the 1977 ETS Invitational Conference, 44-51.

Shoben, E.W. (1978). Differential pass-fail rates in employment testing: Statistical proof under Title VII. Harvard Law Review, 91, 793-813.

Shoben, E.W. (1980). In defense of disparate impact analysis under Title VII: A reply to Dr. Cohn. Indiana Law Journal, 55, 515-536.

Zeisel, J. (1978). Statistics as legal evidence. In International Encyclopedia of Statistics (W.H. Kruskal and J.M. Tanner, Eds.). London: Collier-MacMillan.

4 Annotated Bibliography of Additional Sources

The work of this project drew upon three distinct literatures relevant to the analyses of achievement data in educational court cases. These three literatures are classified according to the professional identities of the authors as: (1) statisticians; (2) legal professionals; and (3) educational researchers. The lack of overlap between the literatures was surprising. The annotated bibliography below is one small attempt to bridge the gaps among these areas.

I Statisticians

Fienberg, S.E., and Straf, M.L. (1982). Statistical assessments as evidence. Journal of the Royal Statistical Society, A, 145, Part 4, 410-421.

The authors discuss the history of statistics in the courts. Statistical evidence was originally considered as hearsay and not admitted as evidence. More recently, methodologies are questioned and attorneys and judges are learning about statistical methods. However, Fienberg and Straf note that "there are no well accepted standards for the preparation, organization and documentation of statistical evidence" (p. 416). When presented as evidence, statistics are regarded as outside the realm of science and inside the realm of justice. This necessitates reinterpretation and reformulation (see p. 420).

Fienberg and Straf present results on the number of cases using statistical terms (e.g. degrees of freedom) and show that there has been a drastic increase since 1960. They also note that a Panel on Statistical Assessments as Evidence in the Courts has been funded by the National Science Foundation. They discuss legal versus scientific inquiry (justice versus truth) and consider two cases (employment discrimination and microwave advertisements). Several interesting observations are capsulized in quotations:

"In examining court records, statisticians might well express surprise as to the methods used and analyses used by statistical witnesses or, more important, as to the methods not used or analyses not done" (p. 415).

"The lawyer may ask the expert to present material in a selective fashion or . . . seek out another statistician whose findings are more acceptable. This type of situation can present the statistician with a moral dilemma" (p. 415).

"If the opposing sides have both introduced expert statistical witnesses, the court almost certainly will be confronted with conflicting testimony" (p. 416).

Fienberg, S.E. and Kadane, J.B. (1983). The presentation of Bayesian statistical analyses in legal proceedings. The Statistician, 34, 88-98.

The authors discuss the use of Bayesian methods in legal cases. They illustrate the approach with a simple example. Then they discuss how some key legal phrases (e.g. "guilty beyond a reasonable doubt") may be interpreted in the Bayesian framework. An actual case is re-examined in some detail.

Kaye, D. (1980). Naked Statistical Evidence. Yale Law Journal, 89, 601-611.

This article is a book review of Quantitative Methods in Law: Studies in the Application of Mathematical Probability and Statistics to Legal Problems, by Finkelstein (1980). It centers on the objective and subjective interpretations of probability that occur in jury selection discrimination cases. The standards of burden of proof and preponderance of evidence are discussed in relation to showing that some event is more likely than chance to occur.

Kaye, D. (1982). The Numbers Game: Statistical proof of discrimination. Michigan Law Review, 80, 837-856.

The article reviews Baldus and Cole (1980), Statistical Proof of Discrimination. Kaye highlights how statistics are used in establishing a prima facie case of discrimination, the use of hypothesis testing, p-values, prediction intervals and Bayesian analyses.

Kaye, D. (1982). Statistical Evidence of Discrimination. Journal of the American Statistical Association, 77, 773-783.

Kaye discusses various statistical techniques for presenting inferential information to a court and illustrates with two jury-selection discrimination cases. Some time is spent discussing the legal meaning of discrimination in various contexts. He notes that standards for handling statistical evidence in courts are still

evolving and notes three issues not yet fully resolved: (a) Is statistical evidence of under-representation sufficient for finding discrimination? (b) When may population analyses be used instead of statistics? (c) What methods should be used to infer that jury selection is independent of, say, race? He discusses each of these in turn.

Statistical techniques considered include p-values, hypothesis testing, prediction (or confidence) intervals, presentation of the likelihood function, and Bayesian analyses. He considers the Bayesian analysis still too controversial for presentation as evidence in court.

The article is followed by comments from several authors.

Downton, F. (1982). Legal probability and statistics. Journal of the Royal Statistical Society-Series A, 145 (4), 395-402.

The author proposes a clarification of legal cases in which statisticians may be involved. The taxonomy is based on the level of statistical sophistication and/or the nature of the role of the statistician in the case. The categories are summarized below.

(i) Non-numerical statistics -- Evidence is verbal and descriptive rather than statistical or mathematical. The statistician is regarded as an expert witness and some attention is given to establishing their credentials and experience.

(ii) Statute statistics -- In these instances, the statistical procedures are incorporated in the law. The statistician and lawyer are in a more client-like relation with the statistician being expected to interpret statistical terminology to non-statisticians.

(iii) Deterministic statistics -- Here the problem is essentially statistical but is not perceived as such by the court. An example is given of an Irish court that regarded actuarial life expectancy tables as non-statistical. The attitude of the court in these cases is one of concern with the facts of a particular case. Hence expected values or general statistical results are discounted.

(iv) Small sample statistics -- In these cases, there are no particular statistical problems. The presentation of statistical evidence plays only a supporting role in the legal system. The probabilistic assumptions for the analysis are a fundamental part of the evidence. However, it is noted that the reluctance of statisticians to attribute "cause" may be interpreted as prevarication by the court.

(v) Applied probability -- In these situations there is usually a small amount of data available for inference, and probabilistic assumptions are less clear. The author gives two examples. He suggests that statisticians avoid cases in this category where the appropriate reference set for the analysis is undefined. Arguments about mathematical assumptions merely obscure the basic problem of inadequacy of information on which calculations should be based.

(vi) Pure Probability -- In this case, probability theory is fundamental to the case (e.g. gambling offenses). However, mathematical equivalence is not the same as legal equivalence. The law is concerned with actual situations and the court prefers specific mathematical results rather than general abstractions (though they may have some bearing).

Meier, P. (1986). Damned liars and expert witnesses. Journal of the American Statistical Association, 81, 394.

Meier suggests that the "needs of the court are not well matched with the usual practice of statistics." Sources of difficulty and suggestions for improvement are outlined. Sources of difficulty include:
Domains of application of statistics in law -- (Sampling, Paternity and fingerprints, Observational data); Inference in discrimination cases; Position of the expert witness; Corrupting influences. His suggestions relate to increased attention to ethical and moral standards.

Fisher, F.M. (1986). Statisticians, econometricians and adversary proceedings. Journal of the American Statistical Association, 81, 394.

This is another discussion of statistician's interaction with the court.

II Legal Professionals

Baldus, D.C., and Cole, J.W. L. (1980). Statistical Proof of Discrimination. New York: McGraw-Hill.

Baldus is professor of law at the University of Iowa; Cole is a consulting statistician. In ten chapters, they discuss a broad range of statistical techniques and typical usage in legal proceedings. The introduction sets the perspective by noting that the validity of a research method (as judged by the court) is assessed in terms of the legal theory of a case. Assessing the "reliability" (i.e., relevance, accuracy) of qualitative information supplied by statisticians requires information about other factors such as the selection process used in collecting data or the characteristics of the original data itself. Much of the book is devoted to discrimination cases. Accordingly, on page 11, the authors give a list of decisions within schools which are commonly subject to claims of discrimination. There was no focus here on achievement testing, growth over time or the problem of determining normal progress.

Chapters 1 and 2 describe various types of court proceedings for disparate treatment cases, disparate impact cases, and some other types of cases. The nature of a "proof" and threats to the "reliability" of quantitative analyses are discussed. It is noted that "an ideal proof will focus on what the substantive law considers relevant and only that . . ." (p. 71).

Chapter 3 discusses measures of actual treatment such as: actual number selected, rates of selection, rates of rejection, rates of representation. (Chapter 2 discussed the use of "actual" measure, "ideal" measure, and "summary" measures [ideal-actual discrepancy]).

Chapter 4 covers relevant comparisons, data collection and analysis. Examples include applicant flow, labor force and definitions of general population. The remaining chapters are briefly summarized below.

Ch. 5: Discussed summary measures (ratios vs. absolute difference); regression and correlation measures; and viewing distributions separately for subgroups (e.g. minority/majority).

Ch. 6: Covers more complex models to take into account job qualifications.

Ch. 7: Covers subgroup comparisons and matching groups.

Ch. 8: Discusses multiple regression and gives an example of investigating a salary dispute adjusting for job qualification.

Ch. 9: Discusses statistical inference--tests and confidence intervals, sample size, the logic of inference and threats to "reliability."

Ch. 10: Is devoted to the interpretation of statistical analyses. Three major considerations are: (1) the risk of error from small samples, and sampling variability; (2) nonprobabilistic threats to validity (e.g. design, measurement error, faulty data); and (3) inferences suggested by other relevant evidence in the case.

Finkelstein, M. (1978). Quantitative Methods in Law, Studies in the Application of Mathematical Probability and Statistics to Legal Problems. New York: The Free Press.

This book is another broad survey, nine chapters, covering cases such as jury discrimination, voting, economic concentration, solvency controls, administrative proceedings (using regression) and compensation for wrongful death. Chapter 1 provides an introduction, discussing the importance of mathematical probability to applications of legal uncertainty. In Chapter 7 the author notes that "the unfamiliarity of regression techniques has undoubtedly impeded their acceptance" (p. 213). An attempt is made to describe and illustrate the technique. The appendix shows a legal application of Bayes' theorem.

Finkelstein, M.O. (1980). The judicial reception of multiple regression studies in race and sex discrimination cases. Columbia Law Review, 80, 737-754.

The author discusses the "statistical war" that has occurred in courts since 1975. He discusses discrimination cases related to promotion, hiring, and back pay. Among the statistical problems considered are inappropriate variables, qualitative factors, data errors and "reverse" regression.

Fisher, F.M. (1980). Multiple regression in legal proceedings. Columbia Law Review, 80, 702-736.

Most of the article is devoted to simple exposition of multiple regression. The author covers uses of multiple regression, the purpose, how it works, estimation, and the assumptions underlying the model. Multiple independent variables and the erroneous inclusion or exclusion of variables is discussed. Another section deals with "goodness-of-fit."

Fisher discusses three types of legal proceedings and whether regression is appropriate to each. He finds that multiple regression is appropriate for wage discrimination cases, of dubious use for investigating anti-trust damages in price fixing cases, and "dangerously misleading" when investigating punishment as a deterrent to crime.

Zeisel, H. (1978). Statistics as legal evidence. In W.H. Kruskal and J.M. Tanur (eds.), International Encyclopedia of Statistics (1118-1121). New York: The Free Press.

This article is divided into three parts. The first part discusses descriptive statistics. Legal precedents are given for using descriptive statistics and survey estimates. Surveys are covered in some detail discussing how they overcome the "hearsay objection" (state-of-mind surveys are an exception) and the fact that surveys sometimes involve experimental design. Surveys must be meticulously executed and documented in order to meet the "peculiar requirements of legal evidence" (i.e. double scrutiny by opposing counsel). Because double scrutiny is "overexacting," one flaw can result in discounting the whole survey as evidence.

Another important point is that some issues of interest are only clarified during the trial. Thus, surveys must be prepared in a general way to anticipate these interests while introducing minimal bias. [Note: the same is probably true for any methodology.]

The second part of the article discusses statistical inference. Its uses are illustrated by anecdotes of parking violations, tax evasion, cheating on a test, and proof of discrimination in jury selection (by comparing actual with expected distribution). It is noted that there are problems with imputing causality from nonexperimental data (one must show that other plausible causes did not in fact cause the effect).

Finally, part three discusses the "evaluation of specific proof." Here the concern is with the reliability of observation and testimony. The court is concerned with "evidentiary power" of, say, a blood test to establish paternity.

Shoben, E.W. (1978). Differential pass-fail rates in employment testing: Statistical Proof under Title VII. Harvard Law Review, 91, 793-813.

Cohn, R.M. (1980a). On the use of statistics in employment discrimination cases. Indiana Law Journal, 55, 493-513.

Shoben, E.W. (1980). In defense of disparate impact analysis under Title VII: A reply to Dr. Cohn. Indiana Law Journal, 55, 515-536.

Cohn R. M. (1980b). Statistical laws and the use of statistics in law: A rejoinder to Professor Shoben. Indiana Law Journal, 55, 537-549.

These articles are all related to employment discrimination and the use of the four-fifths rule in Title VII. The first article, by Shoben (1978) recommends the abandonment of the four-fifths rule and recommends instead a statistical test of the difference between independent proportions to decide whether pass/fail rates, say, are different for blacks and whites. (Appendix describes the Fisher Exact Test.)

Cohn (1980a) more generally discusses "three issues concerning the potential misuse of quantitative information in employment discrimination cases" (p. 493). First, the author argues against the use of inference to indicate adverse impact, on the grounds that he believes an employer's quantitative data should be regarded as population data. It is not a sample from some larger population of interest. Second, Cohn argues that inappropriate measures are frequently used. For example, the ratio of qualified applicants who are hired is of key interest but usually not directly observable. Thus, "disaggregated" measures of the selection process should be used in conjunction with "direct qualitative analyses of the procedures used in the process" (p. 504). Third, the author offers what he believes to be the proper analysis of quantitative data to indicate discrimination. It includes a discussion of controlling for intervening variables. Cohn contends there is no need for lawyers to become statisticians but they must be able to evaluate "qualitative information as probative", to employment discrimination.

Shoben (1980) strongly believes that Cohn has addressed the wrong legal questions. First, she notes "in class actions, the class is typically defined as 'all present and future applicants'" (p. 518). Logic also dictates, she says, that results for a narrowly defined group at a particular time have little or no relevance for long-term policy decisions. Second, she says that Cohn is concerned with passing rates among "qualified" applicants and assumes that selection procedures are valid. But "the selection procedures themselves are at issue if they have the effect of disproportionately excluding a group protected by the Act for <u>whatever</u> reason . . . . Title VII analysis begins with the egalitarian assumption that abilities are equally distributed among groups in our society" (p. 525) except in certain special skill areas (e.g. doctors, lawyers, teachers). Last, she argues that Cohn's use of correlation analysis ignores the Supreme Court formulation of Bonafide Occupational Qualification and the decision that correlations are insufficient proof (p. 533); Cohn's partial correlation analysis assumes that job qualifications have already been established--but they must be litigated.

Cohn (1980b) says Shoben is wrong in considering "all present and future applicants" as a population. "The legally defined group simply cannot be described in quantitative terms" (p. 540). Cohn rejoins that the four-fifths rule was the issue in his paper--not the replacement of validation studies of selection procedures. His examples were geared to show that the 4/5 rule could be violated whether or not there was job relatedness of selection procedures. He argued that the rule was poor because there might be several causes of its violation.

Cohn seems to argue that Shoben misunderstood the intent of his example. Cohn seems to be saying that the courts' efforts to quantify have been off-target. He notes that problems of operationalization of concepts occur in all fields of social inquiry.

III Educational Researchers

Gardner, E. (1982). Some aspects of the use and misuse of standardized aptitude and achievement tests. In Widsor and Gardner (Eds.), Ability Testing II (315-334). Washington: National Academy Press.

Gardner cites some principal misuses of aptitude and achievement tests (p. 323). Among them are failure to report to parents in a meaningful way; misinterpreting measures of changes over a period of time by comparison of results from tests dissimilar in content or norming procedure; use of grade equivalents for profiling achievement tests. In essence, he concludes, "the problems facing measurement today and throughout this decade are primarily political rather than technical in nature" (p. 315).

Findley, W.G. (1979). Civil rights and elementary/secondary school measurement. New Directions for Testing and Measurement, 3, 79-85.

Findley discusses measurement practices used in three types of civil rights and equal protection cases. They are cases involving (1) student group assignments, (2) evaluation of minority programs and (3) the setting of minimum competency standards for graduation. With regard to the first two, he discusses problems with measuring gains by means of grade equivalents and the percentile rank scores of nationally normed tests. Findley describes the use of percentile equivalents of average raw scores for a group coupled with the concept of normal growth as maintaining relative rank over a period of time. An example is described showing shifts in percentile equivalents by ability level.

With regard to competency testing, Findley notes that the two central issues relevant to civil rights are test bias and unequal opportunity to meet standards.

Gable, R.K. and Iwanicki, E.F. (1986). The longitudinal effects of a voluntary school desegregation program on the basic skill progress of participants. Metropolitan Education, No. 1, Spring 1986.

The authors analyze longitudinal data for five years from the implementation of Project Concern. The analyses consisted of t-tests between the experimental and control groups each year. They found no significant differences (with one exception). They concluded that there were no major systematic differences between the Project Concern group and the central group over the five years. However, they argued, this does not imply ineffectiveness of Project Concern.

ERIC

APPENDIX

COMPUTER PROGRAMS FOR ANALYSIS OF STUDENT PROGRESS

1. Listing of Mainframe Program in SAS (PROC MATRIX)
2. Listing of MS-DOS Microcomputer Program in GAUSS

```
//TP9 JOB GLW$KA
// EXEC SAS,OPTIONS='NODATE'
//IN DD DSN=WYL.KA.GLW.SFMRG2,UNIT=DISK,DISP=SHR,
//  VOL=SER=PUB036
//OUT1 DD DSN=WYL.KA.GLW.RTSS.INDSUM,UNIT=DISK,DISP=(NEW,CATLG),
// VOL=SER=PUB036,DCB=(RECFM=FB,LRECL=80,BLKSIZE=6320),
// SPACE=(TRK,(2,1),RLSE)
//SYSIN DD *
*****************************************************************;
*                                                               ;
*              PROGRAM:  TIMEPATH (VERSION 9)                   ;
*                                                               ;
*                         WRITTEN BY                            ;
*                                                               ;
*             JOHN B. WILLETT AND GARY L. WILLIAMSON            ;
*                                                               ;
*     VERSION 9 MODIFIED AND ADAPTED BY GARY L. WILLIAMSON      ;
*                    FROM EARLIER VERSIONS                      ;
*                                                               ;
*                                                               ;
*                                                               ;
*                                                               ;
*                                                               ;
*                                                               ;
*                                                               ;
*   LINEAR TIME-PATHS ARE FIT FOR EACH INDIVIDUAL AND           ;
*   SUMMARY STATISTICS ARE PRINTED FOLLOWED BY A LISTING OF     ;
*   THE ESTIMATE OF RATE OF CHANGE AND R-SQUARE FOR EACH        ;
*   REGRESSION.  THE ANALYSES ARE COMPLETED FOR INDIVIDUALS     ;
*   WHO HAVE SCORES AT ALL TIME POINTS.  (AT THE                ;
*   BEGINNING OF THE PROGRAM, PROC MEANS IS RUN FOR THE ENTIRE  ;
*   DATA SET AND FOR THE REDUCED SET OF INDIVIDUALS HAVING      ;
*   ALL SCORES, FOR THE SAKE OF COMPARISON.)                    ;
*                                                               ;
*                                                               ;
*   CASES WITH HIGH RESIDUAL SUMS OF SQUARES ARE 'TRIMMED',     ;
*   THAT IS DELETED, BEFORE THE CALCULATION OF MAXIMUM          ;
*   LIKELIHOOD ESTIMATES.  (FOULKES-DAVIS' GAMMA AND ITS        ;
*   STANDARD ERROR ARE BASED ON THE FULL DATA.)  THE PROGRAM    ;
*   IS SET TO AUTOMATICALLY TRIM THE TOP 3% OF CASES BASED      ;
*   ON HIGH VALUES OF 'SSRES'.  THE DEFAULT MAY BE CHANGED      ;
*   BY RESETTING THE VALUE OF THE VARIABLE '_TRIM_' IN          ;
*   THE MACRO STATEMENTS AT THE BEGINNING OF THE PROGRAM.       ;
*                                                               ;
*   THIS VERSION OF "TIMEPATHS" ALLOWS FOR THE INCLUSION OF     ;
*   CORRELATES OF CHANGE.  VARIOUS STATISTICS ARE CALCULATED    ;
*   PERTAINING TO SYSTEMATIC INDIVIDUAL DIFFERENCES IN GROWTH.  ;
*                                                               ;
*                                                               ;
*   THE FOLLOWING ANALYSES APPEAR IN ORDER:                     ;
```

```
*                                                                   ;
*   PROC MEANS FOR THE ENTIRE DATA SET                              ;
*   PROC MEANS FOR THE SET OF INDIVIDUALS HAVING SCORES             ;
*      AT ALL TIME POINTS                                           ;
*   PROC MATRIX--THE INDIVIDUAL REGRESSIONS                         ;
*     (STUDENT ID, RATE AND R-SQUARE ARE PRINTED,                   ;
*      AUGMENTED BY ACTUAL RAW SCORE DATA.  IN ADDITION             ;
*      DELTA R-SQUARE IS PRINTED WHERE FITTING A                    ;
*      QUADRATIC SIGNIFICANTLY INCREASES R-SQUARE                   ;
*      (ALPHA=.10)).                                                ;
*   PROC UNIVARIATE ON RATE, R-SQUARE AND SS(RES)                   ;
*   DIAGNOSTICS AND FULL DATA FOR THE 10% OF STUDENTS WITH          ;
*      LOWEST R-SQUARE, THE 10% OF STUDENTS WITH HIGHEST            ;
*      R-SQUARE AND THE 10% OF STUDENTS WITH HIGHEST RESIDUAL       ;
*      SUM OF SQUARES                                               ;
*   CORRELATIONS AMONG THE SCORES AND RATE                          ;
*   PLOTS OF RATE VS SCORE AT EACH TIME                             ;
*   PLOTS OF W VS SELECTED VARIABLES (WHEN CORRELATES OF            ;
*      CHANGE, W,  ARE INCLUDED IN THE ANALYSES)                    ;
*   OBSERVED CORRELATIONS FOR THE TRIMMED DATA                      ;
*   LIST OF CASES DELETED DUE TO HIGH RESIDUAL SUMS OF SQUARES      ;
*   MAXIMUM LIKELIHOOD ESTIMATES (BLOMQVIST), FOULKES-DAVIS         ;
*   ESTIMATES FOR SYSTEMATIC INDIVIDUAL DIFFERENCES IN              ;
*      GROWTH                                                       ;
*                                                                   ;
*                                                                   ;
********************************************************************;


****************************************************************;
*                                                              ;
*   NOTES:                                                     ;
*                                                              ;
*   THE FOLLOWING MACRO DEFINITIONS DEFINE VALUES              ;
*   SPECIFIC TO THIS PARTICULAR EXECUTION OF "TIMEPATHS".      ;
*   IN ORDER TO ALTER THE PROGRAM TO EXECUTE USING             ;
*   DIFFERENT VARIABLES, TIMES, OR TRIM FRACTION, THE          ;
*   APPROPRIATE CHANGES MUST BE MADE IN THESE MACRO            ;
*   DEFINITIONS.  IT MAY ALSO BE NECESSARY TO MODIFY           ;
*   INPUT STATEMENTS.                                          ;
*                                                              ;
****************************************************************;


%LET _XLABEL_=%STR('RTSS9' 'RTSS10' 'RTSS11' 'RTSS12');
%LET _TLABEL_=%STR('9' '10' '11' '12');
%LET _WLABEL_=%STR('W1' 'W2');
%LET _X1_=RTSS9;
%LET _X2_=RTSS10;
%LET _X3_=RTSS11;
```

```
%LET _X4_=RTSS12;
%LET _W1_=W1;
%LET _W2_=W2;
%LET WNAMES=%STR(W1 W2);
%LET _TIME_=%STR(9 10 11 12);
%LET _TRIM_=.03;
%LET _L_=0;
%LET _GAMMA_=1;
%LET _OUT_=1;
%LET TEST=CTBS;
%LET TIMES=GRADE;
%LET T_I=9;
%LET T_F=12;
%LET PLOTLBL=%STR(CTBS READING TOTAL SCALE SCORE);
%LET GRAPH=0;


OPTIONS MISSING=' ';



*************************************************;
*************************************************;
**                                             **;
**          MACRO PROCEDURES                   **;
**                                             **;
*************************************************;
*************************************************;


%MACRO WITH_W;
DATA ALL; INFILE IN; INPUT
#1 ID $ 44-51 SX $ 97 #2 &_X1_ 67-69 #6 &_X2_ 67-69 #10 &_X3_ 67-69
#13 GIFTED $ 53 #14 &_X4_ 67-69 #16;
LENGTH &_X1_ &_X2_ &_X3_ &_X4_ 3;
   IF SX='M' THEN &_W1_=1;
   IF SX='F' THEN &_W1_=0;
   IF SX NE 'M' AND SX NE 'F' THEN &_W1_=.;
   IF GIFTED='1' THEN &_W2_=1; ELSE &_W2_=0;
DROP SX GIFTED;
%MEND WITH_W;




%MACRO W_OUT_W;
DATA ALL; INFILE IN; INPUT
#1 ID $ 44-51 #2 &_X1_ 67-69 #6 &_X2_ 67-69 #10 &_X3_ 67-69
#14 &_X4_ 67-69 #16;
LENGTH &_X1_ &_X2_ &_X3_ &_X4_ 3;
```

```
%MEND W_OUT_W;




%MACRO READDATA;
%IF &_L_ GE 1 %THEN %WITH_W;
%ELSE %W_OUT_W;
%MEND READDATA;




%MACRO ELIM_W;
%IF &_L_=0 %THEN %DO;
        &_W1_=999;
        &_W2_=999;
        DROP &WNAMES;
                    %END;
%MEND ELIM_W;



%MACRO PRT_DLT;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC SORT DATA=DELET; BY SSRES;
PROC PRINT DATA=DELET; ID ID; DROP INT;
     FORMAT RATE RSQ SSRES DEL_R2 F 5.1;
TITLE4 CASES DELETED DUE TO HIGH RESIDUAL SUMS OF SQUARES;
                            %END;
%MEND PRT_DLT;




%MACRO DATAFULL;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC SORT DATA=RESULTS; BY ID;
PROC SORT DATA=TRIMIND; BY ID;
DATA FULL; MERGE RESULTS TRIMIND; BY ID;
IF TRIMIND=0 THEN TRIMD='*';  IF TRIMIND=1 THEN TRIMD='T';
DROP TRIMIND;
                          %END;
%ELSE %DO;
DATA FULL; SET RESULTS;
        %END;
%MEND DATAFULL;
```

```
%MACRO PLOTDAT;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC SORT DATA=FULL; BY DESCENDING TRIMD;
PROC PLOT DATA=FULL; PLOT RATE*(&_X1_ &_X2_ &_X3_ &_X4_)=TRIMD;
TITLE4  ;
                       %END;
%ELSE %DO;
PROC PLOT DATA=FULL; PLOT RATE*(&_X1_ &_X2_ &_X3_ &_X4_);
TITLE4  ;
        %END;
%MEND PLOTDAT;

%MACRO PLOTW;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC PLOT DATA=FULL;
      PLOT (RATE RSQ &_X1_ &_X2_ &_X3_ &_X4_)*(&WNAMES)=TRIMD;
                        %END;
%ELSE %DO;
PROC PLOT DATA=FULL;
      PLOT (RATE RSQ &_X1_ &_X2_ &_X3_ &_X4_)*(&WNAMES);
       %END;
%MEND PLOTW;

%MACRO TRIMCORR;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC CORR NOPROB NOSIMPLE DATA=T_DATA;
      VAR INT RATE &_X1_ &_X2_ &_X3_ &_X4_;
TITLE4 OBSERVED CORRELATIONS ON TRIMMED DATA;
                            %END;
%MEND TRIMCORR;

%MACRO TRMCORW;
%IF &_TRIM_ NE 0 %THEN %DO;
PROC CORR NOPROB NOSIMPLE DATA=T_DATA;
      VAR &WNAMES INT RATE &_X1_ &_X2_ &_X3_ &_X4_;
TITLE4 OBSERVED CORRELATIONS ON TRIMMED DATA;
                            %END;
%MEND TRMCORW;

%MACRO OUTPUT;
```

```
IND_SUM=BETA||SSREG||SSRES||SSTOT||RSQ||
        BETA2||SSREG2||SSRES2||SSTOT2||RSQ2||
        DEL_R2||F||SIGNIF||GAMMA_I||TRIMIND;

CC='INT1' 'RATE' 'SSREG' 'SSRES' 'SSTOT' 'RSQ'
   'INT2' 'B1_HAT' 'B2_HAT' 'SSREG2' 'SSRES2' 'SSTOT2' 'RSQ2'
   'DEL_R2' 'F' 'SIGNIF' 'GAMMA_I' 'TRIMIND';

OUTPUT IND_SUM ROWNAME=ID COLNAME=CC
       OUT=IND_SUM(RENAME=(ROW=ID));
%MEND OUTPUT;



*********************************************;
*********************************************;
**                                         **;
**    END MACRO PROCEDURES                 **;
**                                         **;
*********************************************;
*********************************************;


****************************************;
*                                      ;
*  BEGIN PROGRAM EXECUTION             ;
*                                      ;
****************************************;



%READDATA


TITLE1 STUDY OF STANFORD AND THE SCHOOLS;
TITLE2 WASHINGTON HIGH DATA, CTBS 1980-1983;


PROC MEANS;
TITLE4 DESCRIPTIVE STATISTICS USING ALL CASES;

DATA SCORES; SET ALL;

%ELIM_W

IF &_L_=0 THEN DO;
IF (&_X1_ NE . & &_X2_ NE . & &_X3_ NE . & &_X4_ NE . );
IF (&_X1_ NE 0 & &_X2_ NE 0 & &_X3_ NE 0 & &_X4_ NE 0);
           END;
IF &_L_ GE 1 THEN DO;
```

```
IF (&_X1_ NE . & &_X2_ NE . & &_X3_ NE . & &_X4_ NE . &
   &_W1_ NE . & &_W2_ NE .);
IF (&_X1_ NE 0 & &_X2_ NE 0 & &_X3_ NE 0 & &_X4_ NE 0);
               END;

PROC MEANS;
TITLE4 DESCRIPTIVE STATISTICS USING NON-MISSING, NON-ZERO CASES;

***********************************************************;
*                                                         ;
*  THIS FIRST SECTION CALCULATES QUANTITIES               ;
*  THAT ARE NECESSARY FOR THE LISTINGS AND                ;
*  DESCRIPTIVE STATISTICS PERTAINING TO THE               ;
*  UNTRIMMED DATA SET.                                    ;
*                                                         ;
***********************************************************;

PROC MATRIX FUZZ;
FETCH TEMP ROWNAME = ID;  *READ DATA INTO MATRIX:   NP X (NT+&_L_);
TIME = &_TIME_;          *THE USER MUST GIVE VALUES;
                            *OF TIMEPOINTS HERE       ;

NC=NCOL(TEMP);              *NUMBER OF COLUMNS--WAVES + OMEGAS;

IF &_L_ GE 1 THEN DO;
OMEGA=TEMP(,(NC-&_L_+1):NC); *PEELS OFF THE OMEGA'S--NP X &_L_;
TEMP=TEMP(,1:(NC-&_L_));     *TEMP NOW CONTAINS ONLY THE X'S ;
               END;          * NP X NT;

NT=NC-&_L_;                  *NUMBER OF TIMEPOINTS;

NP=NROW(TEMP);               *NUMBER OF PEOPLE;

IN_TIME=TIME(1,1);        *INITIAL TIME ;

TIME=TIME-J(1,NT,IN_TIME);     *ADJUST TIMES TO START AT ZERO;

T=J(NT,1,1)||TIME';          *THE TIME MATRIX;
T2=J(NT,1,1)||TIME'||(TIME'##2);

X=TEMP';                     *ORIENTS THE DATA MATRIX (NT X NP);
```

```
*****************************************************************;
*                                                               ;
*   CALCULATE REGRESSION ESTIMATES AND R-SQUARE VALUES          ;
*                                                               ;
*****************************************************************;

BETA=INV(T'*T)*(T'*X);
BETA2=INV(T2'*T2)*(T2'*X);

SSREG=VECDIAG(BETA'*T'*X)-(VECDIAG(X'(,+)*X(+,))#/NT);
SSREG2=VECDIAG(BETA2'*T2'*X)-(VECDIAG(X'(,+)*X(+,))#/NT);

SSTOT=VECDIAG(X'*X)-(VECDIAG(X'(,+)*X(+,))#/NT);
SSTOT2=VECDIAG(X'*X)-(VECDIAG(X'(,+)*X(+,))#/NT);

RSQ=((SSREG#/SSTOT)#100); *COL VEC. OF R-SQUARE;
RSQ2=((SSREG2#/SSTOT2)#100);

BETA=BETA';                        *COL. VEC. OF ESTIMATES (NP X 2);
BETA2=BETA2';

DEL_R2=RSQ2-RSQ;

F=ABS(DEL_R2#/((J(NP,1,100)-RSQ2)#/(NT-3)));

SIGNIF=J(NP,1,1)-PROBF(F,1,NT-3);

SSRES=SSTOT-SSREG;    *SSRES IS A VECTOR OF RESIDUAL SUMS OF    ;
                      *SQUARES        (NP X 1)                  ;

SSRES2=SSTOT2-SSREG2;

*****************************************************************;
*                                                               ;
*   CALCULATE FOULKES-DAVIS GAMMA AND SE(GAMMA)                 ;
*                                                               ;
*****************************************************************;

IF &_GAMMA_=1 THEN DO;

T1=MIN(TIME);

T2=MAX(TIME);

PHI=(I(NP)+(J(NP,1,1)*BETA(1:NP,1)')-(BETA(1:NP,1)*J(1,NP,1)))#/

     (I(NP)+(BETA(1:NP,2)*J(1,NP,1))-(J(NP,1,1)*BETA(1:NP,2)'));
```

```
PHI=¬((PHI>=T1)&(PHI<=T2));

PHI=PHI-DIAG(PHI);

GAMMA_I=PHI(1:NP,+)#/(NP-1);       *MATRIX OF GAMMA_I;


GAMMA=SUM(GAMMA_I)#/NP;            *FOULKES-DAVIS GAMMA;

V_GAMMA=SSQ(GAMMA_I-J(NP,1,GAMMA))#/(NP-1);

STER_GAM=SQRT(V_GAMMA#/NP);

FREE PHI;


                  END;
ELSE DO;
     GAMMA_I=J(NP,1,999);
     GAMMA=999;
     STER_GAM=999;
     END;


****************************************************************;
*  THE TIMEPATH ESTIMATES AND R-SQUARES ARE STORED IN A        ;
*  MATRIX CALLED 'RESULT' ALONG WITH THE RAW DATA.             ;
****************************************************************;

IF &_L_ GE 1 THEN
RESULT=OMEGA||BETA(,2)||RSQ||DEL_R2||F||SIGNIF||TEMP;
ELSE
RESULT=BETA(,2)||RSQ||DEL_R2||F||SIGNIF||TEMP;

***************************************************************;
*   CREATE MATRICES CONTAINING TOP AND BOTTOM TEN PERCENT     ;
*   OF CASES BASED ON RSQ                                     ;
***************************************************************;

R=RANK(RESULT(,&_L_+2));   *ASSOCIATE RANKS WITH RSQ VALUES;
TOPRANK=FLOOR(.9#NP);     *CALCULATE NUMBER OF CASES IN TOP 10%  ;
BOTRANK=CEIL(.1#NP);      *CALCULATE NUMBER OF CASES IN BOTTOM 10% ;

TR=LOC(R>=J(NP,1,TOPRANK));  * INDICATOR MATRIX FOR ROWS OF  ;
                             * 'RESULT' THAT HAVE R-SQUARES  ;
                             * IN THE TOP 10%                ;

BR=LOC(R<=J(NP,1,BOTRANK));  * INDICATES ROWS WITH R-SQUARES ;
                             * IN BOTTOM 10%                 ;
```

```
*****************************************************************;
*   SELECT OUT APPROPRIATE ROWS OF 'RESULT' AND THE              ;
*   CORRESPONDING MATRIX OF ID NUMBERS                           ;
*****************************************************************;

TOPTEN=RESULT(TR,);     TOPSTU=ID(TR,);
BOTTEN=RESULT(BR,);     BOTSTU=ID(BR,);

IF &_L_ GE 1 THEN
CC=&_WLABEL_ 'RATE' 'RSQ' 'DEL_R2' 'F' 'SIGNIF'
   &_XLABEL_;
ELSE
CC='RATE' 'RSQ' 'DEL_R2' 'F' 'SIGNIF'
   &_XLABEL_;
OUTPUT TOPTEN COLNAME=CC ROWNAME=TOPSTU OUT=TOP(RENAME=(ROW=ID));
OUTPUT BOTTEN COLNAME=CC ROWNAME=BOTSTU OUT=BOT(RENAME=(ROW=ID));


****************************************************************;
*                                                              ;
*   TOP TEN PERCENT OF RESIDUAL SUMS OF SQUARES                ;
*                                                              ;
****************************************************************;

IF &_L_ GE 1 THEN
RESULT=OMEGA||BETA(,2)||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;
ELSE
RESULT=BETA(,2)||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;

R2=RANK(RESULT(,&_L_+3));     *ASSOCIATE RANKS WITH SSRES VALUES;

TRV=LOC(R2>=J(NP,1,TOPRANK));  *INDICATOR MATRIX FOR ROWS  ;
                               *OF 'RESULT' WITH TOP 10%   ;
                               *OF RESIDUAL SS             ;

TOPTRV=RESULT(TRV,);     *SELECT APPROPRIATE ROWS OF 'RESULT';

TRVSTU=ID(TRV,);         *SELECT CORRESPONDING STUDENT ID'S  ;

IF &_L_ GE 1 THEN
CC1=&_WLABEL_ 'RATE' 'RSQ' 'SSRES' 'DEL_R2' 'F' 'SIGNIF'
    &_XLABEL_;
ELSE
CC1= 'RATE' 'RSQ' 'SSRES' 'DEL_R2' 'F' 'SIGNIF'
     &_XLABEL_;

OUTPUT TOPTRV COLNAME=CC1 ROWNAME=TRVSTU OUT=RESVAR(RENAME=(ROW=ID));


*****************************************************************;
```

```
*                                                                   ;
*   MODIFY R-SQUARE VECTOR TO PRINT ONLY SIGNIFICANT RSQ            ;
*   AND OUTPUT TO 'RESULTS'                                         ;
*                                                                   ;
********************************************************************;

NDEL_R2=DEL_R2#(SIGNIF<=J(NP,1,.10));

IF &_L_ GE 1 THEN DO;
RESULT2=OMEGA||BETA(,2)||RSQ||SSRES||NDEL_R2||TEMP||GAMMA_I;
C =&_WLABEL_ 'RATE' 'RSQ' 'SSRES' 'DEL_R2'
   &_XLABEL_ 'GAMMA_I';
END;

IF &_L_=0 THEN DO;
RESULT2=BETA(,2)||RSQ||SSRES||NDEL_R2||TEMP||GAMMA_I;
C = 'RATE' 'RSQ' 'SSRES' 'DEL_R2'
    &_XLABEL_ 'GAMMA_I';
END;

OUTPUT RESULT2 COLNAME=C ROWNAME=ID OUT=RESULTS(RENAME=(ROW=ID));

***************************************************************;
***************************************************************;
**                                                           *;
**     THIS PART OF THE PROGRAM CALCULATES  QUANTITIES       *;
**     THAT APPEAR ON THE MAXIMUM LIKELIHOOD ESTIMATES       *;
**     SUMMARY PAGE AT THE END OF THE OUTPUT.                *;
**                                                           *;
**     THESE QUANTITIES ARE BASED ON THE TRIMMED             *;
**     DATA.                                                 *;
**                                                           *;
***************************************************************;
***************************************************************;

****************************************************************;
*                                                              ;
*   FIRST:         TRIM OUT CASES WITH HIGHEST RESIDUAL        ;
*   SUMS OF SQUARES (THE FRACTION IS GIVEN BY THE              ;
*   VARIABLE 'TRIM') AND RESET THE BASIC VARIABLE              ;
*   VALUES NEEDED FOR THE ANALYSES ON THE REMAINING            ;
*   DATA.   THE DELETED DATA IS SAVED SO THAT A REFERENCE ;
*   LIST MAY BE PRINTED.                                       ;
*                                                              ;
```

```
****************************************************************;

IF &_TRIM_=0 THEN DO;
     TRIMIND=J(NP,1,0);
     %OUTPUT
     IF &_L_ GE 1 THEN
        RETAIN=OMEGA||BETA||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;
        ELSE
        RETAIN=BETA||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;
     GOTO MLE;
             END;

TRIM=&_TRIM_;   *A CONSTANT INDICATING THE PERCENT OF CASES ;
                *TO BE TRIMMED BECAUSE OF HIGH SSRES--STATE ;
                *IN MACRO AT BEGINNING OF PROGRAM, USING    ;
                *DECIMAL FORM. (I.E., .03 MEANS 3 PERCENT)  ;

TRIMRANK=NP-ROUND(TRIM#NP)+1;  *THE CUT NUMBER--CASES WITH  ;
                               *RANK >= TRIMRANK WILL BE    ;
                               *DELETED                     ;

IF &_L_ GE 1 THEN
DATA=OMEGA||BETA||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;
ELSE
DATA=BETA||RSQ||SSRES||DEL_R2||F||SIGNIF||TEMP;

         * 'DATA' CONTAINS THE BASIC DATA TO BE THINNED ;

RDATA=RANK(DATA(,&_L_+4));  *ASSOCIATE RANKS WITH THE SSRES VARIABLE;

IND1=LOC(RDATA >= J(NP,1,TRIMRANK));  *INDICATOR MATRIX FOR ROWS OF ;
                                      *DATA WITH TRIM FRACTION OF   ;
                                      *HIGHEST SSRES                  ;

IND2=LOC(RDATA < J(NP,1,TRIMRANK));   *INDICATOR FOR ROWS TO BE     ;
                                      *RETAINED                     ;

DELETED=DATA(IND1,);        *SELECT HIGHEST SSRES CASES ;
THINSTU=ID(IND1,);          *SELECT CORRESPONDING STUDENT ID'S ;
STAYSTU=ID(IND2,);

IF &_L_ GE 1 THEN
CL1=&_WLABEL_ 'INT' 'RATE' 'RSQ' 'SSRES' 'DEL_R2' 'F' 'SIGNIF'
     &_XLABEL_;
ELSE
CL1= 'INT' 'RATE' 'RSQ' 'SSRES' 'DEL_R2' 'F' 'SIGNIF'
     &_XLABEL_;

OUTPUT DELETED COLNAME=CL1 ROWNAME=THINSTU
     OUT=DELET(RENAME=(ROW=ID));
```

```
RETAIN=DATA(IND2,);   *SELECT CASES FOR 'TRIMMED' ANALYSIS;

OUTPUT RETAIN COLNAME=CL1 ROWNAME=STAYSTU OUT=T_DATA;

   ***********************************************;
   *                                             ;
   *  CREATE A PLOTTING VARIABLE TO IDENTIFY     ;
   *  TRIMMED CASES IN PLOTS                     ;
   *                                             ;
   ***********************************************;


TRIMIND=J(NP,1,0);

TRIMIND1=J(NP,1,1);

TRIMIND(IND1,)=TRIMIND1(IND1,);

CNAME='TRIMIND';

OUTPUT TRIMIND ROWNAME=ID COLNAME=CNAME
       OUT=TRIMIND(RENAME=(ROW=ID));


*************************************************************;
*                                                           ;
*  OUTPUT INDIVIDUAL SUMMARY FILE                           ;
*                                                           ;
*************************************************************;

%OUTPUT



   ***********************************************;
   *                                             ;
   *  RESET KEY DATA VARIABLES TO VALUES TO BE   ;
   *  USED FOR THE TRIMMED ANALYSIS              ;
   *                                             ;
   ***********************************************;

BETA=RETAIN(,&_L_+1:&_L_+2);
RSQ=RETAIN(,&_L_+3);
SSRES=RETAIN(,&_L_+4);
TEMP=RETAIN(,&_L_+8:&_L_+7+NT);
IF &_L_ GE 1 THEN OMEGA=RETAIN(,1:&_L_);
NP=NROW(TEMP);              *NUMBER OF CASES LEFT AFTER TRIMMING;
```

```
FREE DATA RDATA IND1 IND2 DELETED THINSTU ;

MLE:;***********************************************************;
*                                                              ;
*   INTERMEDIATE STATISTICAL QUANTITIES NECESSARY FOR          ;
*   THE CALCULATION OF BLOMQVIST ESTIMATES                     ;
*                                                              ;
*   NOTE THAT THE TIME VECTOR IS RECENTERED SO THAT IT STARTS ;
*   AT TIME ZERO.  THIS IS DONE TO FACILITATE THE APPLICATION ;
*   OF BLOMQVIST'S FORMULAE.  SUBSEQUENT ESTIMATES FOR THE     ;
*   COVARIANCE MATRIX FOR ESTIMATED INITIAL STATUS AND RATE    ;
*   OF CHANGE INCLUDE ADJUSTMENT BACK TO THE ORIGINAL TIME     ;
*   SCALE (WHERE GRADE NINE IS DEFINED AS INITIAL STATUS).     ;
*   T_ZERO IS SIMILARLY ADJUSTED LATER.                        ;
****************************************************************;

TIMEPTS=T(,2);  *ADJUST TIMES TO START AT 0;

TBAR=TIMEPTS(+,1)#/NT;          *MEAN TIME;

TSQ=TIMEPTS##2;
TSQ=TSQ(+,1);
SST=TSQ-(NT#TBAR#TBAR);         *SUM OF SQUARES FOR TIME;

BMEAN=BETA(+,)#/NP;          *ROW VECTOR OF MEAN EST'D. IN. ST. & RATE;

BETA=BETA-J(NP,1,1)*BMEAN;  *DEVIATION MATRIX FOR PARAMETER ESTS.;
SSBETA=BETA'*BETA;
COVBETA=SSBETA#/NP;          *MLE OF THE COVARIANCE MATRIX OF         ;
                             *ESTIMATED INITIAL STATUS (GRADE NINE) ;
                             *AND RATE OF CHANGE                      ;

****************************************************************;
*                                                              ;
*   BLOMQVIST ESTIMATES OF COVARIANCE MATIX OF INITIAL         ;
*   STATUS AND RATE OF CHANGE                                  ;
*                                                              ;
****************************************************************;

ERRVAR=SSRES(+,)#/(NP#(NT-2));   *POOLED ERROR VARIANCE;
```

```
VARINST=COVBETA(1,1)-((ERRVAR#TSQ)#/(SST#NT));

V_TRATE=COVBETA(2,2)-(ERRVAR#/SST);

COVISTR=COVBETA(2,1)+(ERRVAR#TBAR#/SST);



*****************************************************************;
*                                                               ;
*   ESTIMATION OF T-ZERO, VARIANCE OF KSI(TO), AND THE          ;
*   SCALING FACTOR.                                             ;
*                                                               ;
*****************************************************************;

T_ZERO=-(COVISTR#/V_TRATE);

V_KSI_TO=((VARINST#V_TRATE)-(COVISTR##2))#/V_TRATE;

SCAL_FAC=SQRT(V_KSI_TO#/V_TRATE);

SEM_RATE=SQRT(COVBETA(2,2)-V_TRATE);


*****************************************************************;
*                                                               ;
*   ESTIMATION OF THE VARIANCE OF KSI(T) AND THE                ;
*   COVARIANCE OF KSI(T) AND T_RATE.                            ;
*                                                               ;
*****************************************************************;

T_TZERO=TIME'-J(NT,1,T_ZERO);

C_KSITR =J(NT,1,V_TRATE)#T_TZERO;

V_KSI_T=J(NT,1,V_KSI_TO)+(J(NT,1,V_TRATE)#(T_TZERO##2));   *NT X 1;

*****************************************************************;
*                                                               ;
*   ESTIMATION OF THE REGRESSION COEFF AND CORRELATION OF       ;
*   T_RATE ON KSI(T).                                           ;
*                                                               ;
*****************************************************************;

B_TRKSI=C_KSITR#/V_KSI_T;

R_TRKSI=C_KSITR#/SQRT(J(NT,1,V_TRATE)#V_KSI_T);


```

```
*****************************************************************;
*                                                               ;
*   FINDING OBSERVED VARIANCE AND BETWEEN-WAVE CORRELATION      ;
*   MATRIX.                                                     ;
*                                                               ;
*****************************************************************;

X=TEMP;
XBAR=X(+,)#/NP;
XDEV=X-J(NP,1,1)*XBAR;
SSX=XDEV'*XDEV;                 * DIMENSIONS  (NT X NT);
COV_X=SSX#/(NP);
VAR_X=VECDIAG(COV_X);                 *NT X 1;
STD_X=SQRT(VAR_X);
CORR_X=COV_X#/(STD_X*STD_X');


*****************************************************************;
*                                                               ;
*   ESTIMATION OF RELIABILITY OF TRATE-HAT AND OF X(I).         ;
*                                                               ;
*****************************************************************;

REL_TR=V_TRATE#/COVBETA(2,2);


REL_XT=V_KSI_T#/VAR_X;    *BASED ON THE SAMPLE--NT X 1 ;

R_KSI=CORR_X#/SQRT(REL_XT*REL_XT');  *TRUE SCORE CORR. MATRIX;

R_KSI=R_KSI-DIAG(R_KSI)+I(NT);   *INSERT 1'S ON THE DIAGONAL;



*****************************************************************;
*                                                               ;
*   STORING THE BLOMQVIST ESTIMATES                             ;
*                                                               ;
*****************************************************************;

T_ZERO_S=T_ZERO + IN_TIME;

R1=T_ZERO_S//SCAL_FAC//V_TRATE//V_KSI_TO//REL_TR//SEM_RATE;

R1A=GAMMA//STER_GAM;

R2=R_KSI;

R3=V_KSI_T'//VAR_X'//REL_XT'//B_TRKSI'//R_TRKSI';
```

```
OUTPUT R1 OUT=R1;
OUTPUT R1A OUT=R1A;
OUTPUT R2 OUT=R2;
OUTPUT R3 OUT=R3;



*************************************************************;
*                                                           ;
*   COMPUTE STATISTICS RELATED TO SYSTEMATIC INDIVIDUAL     ;
*   DIFFERENCES IN GROWTH.                                  ;
*                                                           ;
*************************************************************;

IF &_L_=0 THEN GOTO PRT;

     ************************************************;
     *                                               
     *  ADDITIONAL COVARIANCE ESTIMATES INVOLVING    ;
     *                   OMEGA                       ;
     *                                               ;
     ************************************************;


D_W=(OMEGA - (J(NP,1)*(OMEGA(+,)#/NP)));      *OMEGA DEVIATIONS ;
                                              *NP X &_L_;

PV = BETA||D_W;     *DEVIATION MATRIX OF PARAMETER VALUES   ;
                    * KSI(9), RATE, OMEGA--NP X (2+&_L_)      ;

C_PV=(PV'*PV)#/NP;  *MLE OF COVARIANCE MATRIX OF PV;

V_W=VECDIAG(C_PV)(3:(2+&_L_),);        *ESTIMATED VARIANCE OF OMEGA'S;
                                       *&_L_ X 1;

C_TH_W=C_PV(3:(2+&_L_),2);    *ESTIMATED COVARIANCE(RATE,OMEGA) ;
                              *&_L_ X 1 VECTOR;

C_KSO_W=C_PV(3:(2+&_L_),1);
                    *ESTIMATED COVARIANCE(INITIAL STATUS,OMEGA) ;
                    *&_L_ X 1;

C_KSTO_W=C_KSO_W - ((C_TH_W#COVISTR)#/V_TRATE);

                    *ESTIMATED COVARIANCE(KSI_TO,OMEGA) ;
                    *&_L_ X 1;


      ************************************************;
```

ERIC

```
       *                                                  ;
       *  COMPUTE REQUIRED CORRELATION AND REGRESSION     ;
       *                 ESTIMATES                        ;
       *                                                  ;
       ***************************************************;

*ALL QUANTITIES IN THIS SECTION ARE VECTORS--&_L_ X 1;

R_TH_W=C_TH_W#/SQRT(V_TRATE#V_W);     *CORRELATION(RATE,OMEGA) ;

B_TH_W=C_TH_W#/V_W;        *REGRESSION COEFF.(RATE,OMEGA)        ;

R_KSTO_W=C_KSTO_W#/SQRT(V_KSI_TO#V_W);

                           *CORRELATION(KSI_TO,OMEGA)            ;

B_KSTO_W=C_KSTO_W#/V_W;    *REGRESSION COEFF.(KSI_TO,OMEGA)      ;

T_U=J(&_L_,1,T_ZERO) + ((V_KSI_TO#C_TH_W)#/(V_TRATE#C_KSTO_W));

                               *T-UPPER;

T_L=J(&_L_,1,T_ZERO) - (C_KSTO_W#/C_TH_W);                *T-LOWER;


       ***************************************************;
       *                                                  ;
       *  COMPUTE WAVE BY WAVE CORRELATION AND            ;
       *  REGRESSION ESTIMATES                            ;
       *                                                  ;
       ***************************************************;

*ALL QUANTITIES IN THIS SECTION ARE MATRICES--&_L_ X NT;
*EACH ROW CORRESPONDS TO A DIFFERENT W;


C_X_W=(D_W'*XDEV)#/NP;

R_X_W=C_X_W#/SQRT((J(&_L_,1)*VAR_X')#(V_W*J(1,NT)));

                            *CORR(X,OMEGA) ;

B_X_W=C_X_W#/(V_W*J(1,NT));              *REGR.COEFF(X,OMEGA);

B_KSI_W=(B_KSTO_W*J(1,NT))+((B_TH_W*J(1,NT))#
                              (J(&_L_,1)*T_TZERO'));

                               *REGR.COEFF(KST,OMEGA);

```

```
R_KSI_W=R_X_W#/SQRT(J(&_L_,1)*REL_XT');

                          *BASED ON RELIABILITY ESTIMATE   ;



*****************************************************************;
*                                                               ;
*   STORING THE ESTIMATES RELATED TO SYSTEMATIC INDIVIDUAL     ;
*   DIFFERENCES IN GROWTH                                       ;
*                                                               ;
*****************************************************************;


T_U_S=T_U + J(&_L_,1)#IN_TIME;

T_L_S=T_L + J(&_L_,1)#IN_TIME;


OMEGA1=R_TH_W'//B_TH_W'//R_KSTO_W'//B_KSTO_W'//T_U_S'//T_L_S';

                  * 6 X &_L_  ;

OMEGA2=R_X_W//R_KSI_W;

                        * (2 X &_L_) X NT  ;

OMEGA3=B_X_W//B_KSI_W;          *  (2 X &_L_) X NT  ;

OUTPUT OMEGA1 OUT=OMEGA1;
OUTPUT OMEGA2 OUT=OMEGA2;
OUTPUT OMEGA3 OUT=OMEGA3;


PRT:;*******************************************************;
*                                                           ;
*   NOW THE PRINTING BEGINS.....                            ;
*                                                           ;
************************************************************;


%DATAFULL


SCORES='          ';
IF DEL_R2=0 THEN DEL_R2=.;
LABEL SCORES=' ' ID=ID;
```

```
%MACRO PRT_WW;
PROC PRINT LABEL SPLIT="*"; ID ID;
     VAR &WNAMES RATE RSQ SCORES DEL_R2 SCORES
                                  &_X1_ &_X2_ &_X3_ &_X4_;
     FORMAT RATE RSQ DEL_R2  5.1;
     TITLE4 FITS TO INDIVIDUAL TIMEPATHS;
PROC UNIVARIATE DATA=RESULTS PLOT NORMAL; ID ID ;
     VAR RATE RSQ SSRES GAMMA_I;
     TITLE4  ;
*************************************************;
*  SORT BY RSQ AND PRINT OUT TOP AND BOTTOM 10% ;
*************************************************;
PROC SORT DATA=BOT; BY RSQ;
PROC PRINT DATA=BOT; ID ID; FORMAT RATE RSQ DEL_R2 F 5.1;
TITLE4 FITS WITH LOWEST R-SQUARE;
PROC SORT DATA=TOP; BY RSQ;
PROC PRINT DATA=TOP; ID ID; FORMAT RATE RSQ DEL_R2 F 5.1;
TITLE4 FITS WITH HIGHEST R-SQUARE;
************************************************************;
*                                                          ;
*   PRINT TOP 10%  SSRES                                   ;
*                                                          ;
************************************************************;
PROC SORT DATA=RESVAR; BY SSRES;
PROC PRINT DATA=RESVAR; ID ID;
     FORMAT RATE RSQ SSRES DEL_R2 F 5.1;
TITLE4 FITS WITH HIGHEST RESIDUAL SUMS OF SQUARES;
*************************************************;
*                                               ;
*   CORRELATIONS OF OBSERVED SCORES AND RATE    ;
*                                               ;
*************************************************;
PROC CORR NOPROB NOSIMPLE DATA=RESULTS;
     VAR &_X1_ &_X2_ &_X3_ &_X4_ &WNAMES RATE;
     TITLE4 OBSERVED CORRELATIONS;
%PLOTDAT
```

```
%PLOTW

PROC SORT DATA=FULL; BY ID;

     * MUST SORT THE DATA BACK INTO ID ORDER AFTER PLOTS;

%TRMCORW


%MEND PRT_WW;

%MACRO PRT_WOW;

PROC PRINT LABEL SPLIT="; ID ID;
     VAR RATE RSQ SCORES DEL_R2 SCORES &_X1_ &_X2_ &_X3_ &_X4_;
     FORMAT RATE RSQ DEL_R2  5.1;
     TITLE4 FITS TO INDIVIDUAL TIMEPATHS;
PROC UNIVARIATE DATA=RESULTS PLOT NORMAL; ID ID ;
     VAR RATE RSQ SSRES GAMMA_I;
     TITLE4  ;

*************************************************;
*  SORT BY RSQ AND PRINT OUT TOP AND BOTTOM 10% ;
*************************************************;

PROC SORT DATA=BOT; BY RSQ;
PROC PRINT DATA=BOT; ID ID; FORMAT RATE RSQ DEL_R2 F 5.1;
TITLE4 FITS WITH LOWEST R-SQUARE;
PROC SORT DATA=TOP; BY RSQ;
PROC PRINT DATA=TOP; ID ID; FORMAT RATE RSQ DEL_R2 F 5.1;
TITLE4 FITS WITH HIGHEST R-SQUARE;


************************************************************;
*                                                           ;
*   PRINT TOP 10%  SSRES                                    ;
*                                                           ;
************************************************************;

PROC SORT DATA=RESVAR; BY SSRES;

PROC PRINT DATA=RESVAR; ID ID;
     FORMAT RATE RSQ SSRES DEL_R2 F 5.1;

TITLE4 FITS WITH HIGHEST RESIDUAL SUMS OF SQUARES;


**************************************************;
*                                                 ;
```

```
*   CORRELATIONS OF OBSERVED SCORES AND RATE       ;
*                                                  ;
***************************************************;


PROC CORR NOPROB NOSIMPLE DATA=RESULTS;
     VAR &_X1_ &_X2_ &_X3_ &_X4_  RATE;
     TITLE4 OBSERVED CORRELATIONS;

%PLOTDAT

PROC SORT DATA=FULL; BY ID;

          *MUST SORT THE DATA BACK INTO ID ORDER AFTER PLOTS;

%TRIMCORR



%MEND PRT_WOW;


%MACRO PRINT;

%IF &_L_ GE 1 %THEN %PRT_WW;

%ELSE %PRT_WOW;

%MEND PRINT;



%PRINT


%PRT_DLT

*************************************************************;
*                                                           ;
*   PRINT MAXIMUM LIKELIHOOD ESTIMATES                      ;
*                                                           ;
*   (ALL QUANTITIES EXCEPT GAMMA AND STER_GAM ARE BASED     ;
*    ON THE TRIMMED DATA)                                   ;
*                                                           ;
*************************************************************;

PROC MATRIX FUZZ;
TITLE4 &_TRIM_ OF CASES TRIMMED;
TITLE5 MAXIMUM LIKELIHOOD ESTIMATES (BLOMQVIST, 1977);
```

```
FETCH FULLDATA DATA=R1A;  C1=' ';
R= 'GAMMA' 'STER_GAM';
PRINT FULLDATA COLNAME=C1 ROWNAME=R FORMAT=8.3;

FETCH TRIMDATA DATA=R1;  C1= ' ';
R= 'T_ZERO' 'SCAL_FAC' 'V_TRATE' 'V_KSI_T0' 'REL_TR' 'SEM_RATE';
PRINT TRIMDATA COLNAME=C1 ROWNAME=R FORMAT=8.3;

FETCH KSI_CORR DATA=R2;   C2= &_TLABEL_;
PRINT KSI_CORR COLNAME=C2 ROWNAME=C2 FORMAT=8.3;

FETCH _ DATA=R3;
R2= 'V_KSI_T' 'V_X_T' 'REL_XT' 'B_TRKSI' 'R_TRKSI' ;
PRINT _ COLNAME=C2 ROWNAME=R2 FORMAT=8.3;

*******************************************************************;
*                                                                 ;
*   PRINT ESTIMATES RELATED TO SYSTEMATIC INDIVIDUAL              ;
*   DIFFERENCES IN GROWTH                                         ;
*                                                                 ;
*   (ALL QUANTITIES BASED ON TRIMMED DATA)                        ;
*                                                                 ;
*******************************************************************;

%MACRO OMEGA;

%IF &_L_ GE 1 %THEN %DO;

PROC MATRIX FUZZ;
TITLE4 &_TRIM_ OF CASES TRIMMED;
TITLE5 SYSTEMATIC INDIVIDUAL DIFFERENCES IN GROWTH;
FETCH _ DATA=OMEGA1;    C1=&_WLABEL_;
R1='R_TH_W' 'B_TH_W' 'R_KST0_W' 'B_KST0_W' 'T_U' 'T_L';
PRINT _ COLNAME=C1 ROWNAME=R1 FORMAT=8.3;

FETCH CORREL DATA=OMEGA2;   C2= &_TLABEL_;
R2='R_X_W' 'R_KSI_W' ;
DO I=1 TO &_L_;
    A=I||&_L_+I||(2*&_L_)+I||(3*&_L_)+I;
    CORR=CORREL(A,);
    PRINT CORR COLNAME=C2 ROWNAME=R2 FORMAT=8.3;
END;

FETCH REGRESS DATA=OMEGA3;
R3='B_X_W' 'B_KSI_W';
DO I=1 TO &_L_;
```

```
      A=I||&_L_+I;
      REGR=REGRESS(A,);
      PRINT REGR COLNAME=C2 ROWNAME=R3 FORMAT=8.3;
   END;
                     %END;
   %MEND OMEGA;
   %OMEGA
   %MACRO FILEOUT;
   %IF &_OUT_=1 %THEN %DO;
   DATA; SET IND_SUM; FILE OUT1;
   PUT ID $ 1-9 (INT1--RSQ) (F8.3)/ @10
                (INT2--RSQ2) (F8.3)/@10
                (DEL_R2--TRIMIND) (F8.3);
                      %END;
   %MEND FILEOUT;
   %FILEOUT
   %MACRO SASGRAPH;
   %IF &GRAPH=1 %THEN %DO;
      DATA GRAFDAT;
         SET IND_SUM;  KEEP ID SCORE T;
          GOPTIONS DEVICE=VEP12FF;
            SCORE=INT1; T=&T_I; OUTPUT;
            SCORE=INT1 + (&T_F-&T_I)*RATE; T=&T_F; OUTPUT;
         LABEL SCORE=&TEST T=&TIMES;
         TITLE .H=3 .F=TRIPLEX &PLOTLBL;
      PROC SORT DATA=GRAFDAT; BY ID;
      PROC GPLOT DATA=GRAFDAT;
      PLOT SCORE*T=ID/CTEXT=BLACK NOLEGEND;
      SYMBOL C=BLACK I=JOIN R=130;
   %END;
   %MEND SASGRAPH;
   %SASGRAPH
```

```
@          Time Path Program

           Adapted from TIMEPATH (version 8) by J.B. Willett and
           Gary L. Williamson, originally translated into GAUSS
           by G. Ghandour.  GAUSS version modified and enhanced by
           Gary L. Williamson, Fall 1986                                  @

new ,25000;
loadp dstatx medianx selif delif recode savextod;

#include menu;

@procedure to calculate quantiles @

proc 1=quantile(x,q);
local k, n, qnt, i, v, j, g, p, svx, indx, sv;

k=cols(x); n=rows(x); qnt=zeros(k,1); p=q./100;
j=trunc((n+1).*p);
g=((n+1).*p) - j;

indx=seqa(1,1,n);
i=1;
do until i > k;
  v=submat(x,0,i);
  svx=sortc((v~indx),1);
  sv=submat(svx,0,1);
  if (j+1 <= n) and (j >= 1);
  qnt[i,1]=((1-g).*sv[j,1]) + g.*sv[j+1,1];
  elseif (j == n);
  qnt[i,1]=sv[j,1];
  elseif (j == 0);
  qnt[i,1]=sv[j+1,1];
  endif;
i=i+1;
endo;
retp(qnt);
endp;

@end of procedure @

output file=^hardcopy reset;
nc=nt+n1;

if d == 1;
load data[np,nc+1]=^infile;
elseif d == 2;
open f1=^infile;
data=readr(f1,np);
endif;

@**********************************************************************@
@ The following section must be manually edited prior to running the @
@ program, in order to accomodate your particular set of data        @
@**********************************************************************@

@ create labels for printing--these must be edited to accomodate the
desired spacing   @
```

```
stlbl1="    rate   rsq      d_rsq        ";
stlbl2="    rate   rsq   ssres      d_rsq       ";
xlbl="  x1      x2      x3      x4";
xlbl2="    x1      x2       x3       x4";
wlbl=" w";
wlbl2="    w   ";
tlbl=" ";

@ if a GAUSS data set is being created, the vector, varnames,   @
@ should be edited to contain the proper number of x's and w's @

if outfile == 1;
if (nl >= 1 and c_of_c >= 1);
let varnames=id x1 x2 x3 x4 w rate rsq ssres gamma_i;
elseif (nl == 0 or c_of_c == 0);
let varnames=id x1 x2 x3 x4 rate rsq ssres gamma_i;
endif;
endif;


@Create print formats--these must be edited whenever the number of
 significant digits in the x's is different from the previous run.

 Edit the first set of masks and fmts if correlates of change are
 included in this run.  Edit the second set of masks and fmts if no
 w's are used. @

if (nl >= 1 and c_of_c > 0);
let mask1[1,9]= 1 1 1 1 1 1 1 1 1 ;
let mask2[1,10]= 1 1 1 1 1 1 1 1 1 1;
let fmt1[9,3]=
   "*.*lf " 7 0
   "*.*lf " 1 0
   "*.*lf " 7 3
   "*.*lf " 5 1
   "*.*lf " 10 1
   "*.*lf " 10 2
   "*.*lf " 6 2
   "*.*lf " 6 2
   "*.*lf " 6 2  ;
let fmt2[10,3]=
   "*.*lf " 7 0
   "*.*lf " 1 0
   "*.*lf " 7 3
   "*.*lf " 5 1
   "*.*lf " 7 2
   "*.*lf " 10 1
   "*.*lf " 10 2
   "*.*lf " 6 2
   "*.*lf " 6 2
   "*.*lf " 6 2  ;
endif;
if (nl == 0 or c_of_c <= 0);
let mask1[1,8]= 1 1 1 1 1 1 1 1 ;
let mask2[1,9]= 1 1 1 1 1 1 1 1 1;
let fmt1[8,3]=
   "*.*lf " 7 0
   "*.*lf " 7 3
   "*.*lf " 5 1
   "*.*lf " 10 1
   "*.*lf " 10 2
   "*.*lf " 6 2
   "*.*lf " 6 2
   "*.*lf " 6 2  ;
let fmt2[9,3]=
```

```
   "*.*lf " 7 3
   "*.*lf " 5 1
   "*.*lf " 7 2
   "*.*lf " 10 1
   "*.*lf " 10 2
   "*.*lf " 6 2
   "*.*lf " 6 2
   "*.*lf " 6 2  ;
endif;

@*******************************************************************@
@  end manual changes                                               @
@*******************************************************************@


@ peel off observed scores, background variables, produce descriptive @
@ statistics                                                          @

id=data[.,1];
tmp=data[.,2:nc+1];
clear data;
"Time Path on " namedata; " "; " " ; " ";
dstatx(tmp);   output on;  "\f";


@ Partition data matrix into timepaths and correlates              @

x=tmp[.,1:nt];
if nl >= 1;
   omg=tmp[.,nt+1:nc];
endif;
clear tmp;

@ adjust time to start at 0                                          @

i_tm=tm[1,1];
tm=tm-i_tm*ones(nt,1);

@ Create the time design matrices                                    @

t_1=ones(nt,1)~tm;
t_2=t_1~tm^2;

@ Re-orient the time path matrix                                     @

x=x';

@ Calculate regression estimates and r-squares                       @

beta_1=solpd(t_1'x,t_1't_1);
beta_2=solpd(t_2'x,t_2't_2);



crf=sumc(x).*sumc(x)/nt;

 idx=1;
 ssreg_1=zeros(np,1);
 ssreg_2=zeros(np,1);
 sstot=zeros(np,1);
do while idx <= np;
 ssreg_1[idx,1]=beta_1[.,idx]'*t_1'*x[.,idx];
 ssreg_2[idx,1]=beta_2[.,idx]'*t_2'*x[.,idx];
 sstot[idx,.]=x[.,idx]'*x[.,idx];
 idx=idx+1;
```

```
ssreg_1=ssreg_1-crf;
ssreg_2=ssreg_2-crf;

sstot=sstot-crf;

rsq_1=ssreg_1./sstot;
rsq_2=ssreg_2./sstot;

d_rsq=rsq_2-rsq_1;

f_v=abs(d_rsq./(ones(np,1)-rsq_2))/(nt-3);
f_s=cdffc(f_v,ones(np,1),(nt-3)*ones(np,1));

ssres_1=sstot-ssreg_1;
ssres_2=sstot-ssreg_2;

clear crf, beta_2, f_s, f_v, rsq_2, ssreg_2, ssres_2, sstot;

@output fits to individual regressions@

namedata; " "; " ";
format /m1 /rd 7,2;
"Fits to individual regressions";
" ";
if (nl >= 1 and c_of_c >= 1);
idlbl$+wlbl$+stlbl1$+xlbl;
" ";
result=id~omg~beta_1[2,.]'~100.*rsq_1~100.*d_rsq~x';
p=printfm(result,mask1,fmt1); "\f";
else;
idlbl$+stlbl1$+xlbl;
" ";
result=id~beta_1[2,.]'~100.*rsq_1~100.*d_rsq~x';
p=printfm(result,mask1,fmt1); "\f";
endif;

@select top and bottom 10% of cases based on rsq    @

sortrslt=sortc(result,3);
botcut=ceil(.1.*np);
topcut=floor(.9.*np);
botten=sortrslt[1:botcut,.];
topten=sortrslt[topcut:np,.];

@print top and bottom 10% based on rsq    @

namedata; " "; " ";
"Fits with lowest R-square";  " ";
if (nl >= 1 and c_of_c >= 1);
idlbl$+wlbl$+stlbl1$+xlbl;
else;
idlbl$+stlbl1$+xlbl;
endif;
p=printfm(botten,mask1,fmt1); " ";  "\f";

namedata; " "; " ";
"Fits with highest R-square";  " ";
if (nl >=1 and c_of_c >= 1);
idlbl$+wlbl$+stlbl1$+xlbl;
else;
idlbl$+stlbl1$+xlbl;
endif;
p=printfm(topten,mask1,fmt1);  "\f";

ear botten, topten;
```


ERIC

```
@Find fits with highest residual sums of squares @

clear result, sortrslt;
if (nl >= 1 and c_of_c >=1);
result=id~omg~beta_1[2,.]'~100.*rsq_1~ssres_1~100.*d_rsq~x';
else;
result=id~beta_1[2,.]'~100.*rsq_1~ssres_1~100.*d_rsq~x';
endif;
sortrslt=sortc(result,4);
topres=sortrslt[topcut:np,.]; " ";

@print top 10% of cases with high SSRES @

namedata; " "; " ";
"Fits with highest residual sums of squares"; " ";
if (nl >= 1 and c_of_c >= 1);
idlbl$+wlbl$+stlbl2$+xlbl;
else;
idlbl$+stlbl2$+xlbl;
endif;
p=printfm(topres,mask2,fmt2); "\f";

clear topres, sortrslt;

@calculate observed correlations @

if nl == 0;
m=x'~beta_1[2,.]';
elseif nl > 0;
m=x'~omg~beta_1[2,.]';
endif;
m_bar=meanc(m);
m_dev=m-ones(np,1)*m_bar';
cov_m=(m_dev'*m_dev)./np;
sd_m=sqrt(diag(cov_m));
corr_m=cov_m./(sd_m*sd_m');
if c_of_c <= 0;
corr_m=(corr_m[1:nt,1:nt]~corr_m[1:nt,nc+1])|
        (corr_m[nc+1,1:nt]~corr_m[nc+1,nc+1]);
endif;

clear m, m_bar, m_dev, cov_m, sd_m;

@print correlations between waves, omega and w @

namedata; " "; " ";
" ";
"Observed correlations, untrimmed data ";
" ";
if (nl >= 1 and c_of_c >= 1);
xlbl2$+wlbl2$+"      rate";
else;
xlbl2$+"      rate";
endif;
format /m1 /rd 7,3;
corr_m;  "\f";


@ Calculate Foulkes-Davis gamma and S.E.(gamma)                      @

 idx=1;
 phi=zeros(np,1);
 gam_i=zeros(np,1);
   while idx <= np;
 phi=beta_1[2,idx]*ones(np,1)-beta_1[2,.]';
 .=(phi   == 0).
```

```
 phi=recode(phi,e,v);
 phi=(beta_1[1,.]'-beta_1[1,idx]*ones(np,1))./phi;
 phi=.not((phi.>=tm[1,1]).and(phi.<=tm[nt,1]));
 v=0;
 phi=recode(phi,e,v);
 gam_i[idx,1]=sumc(phi)/(np-1);
 idx=idx+1;
endo;
clear phi, e;

gam_fd=meanc(gam_i);
se_gam=stdc(gam_i)./sqrt(np);
x=x';
beta_1=beta_1';

@descriptive statistics on rate, rsq, ssres and gamma@

tmp=beta_1[.,2]~100*rsq_1~ssres_1~gam_i;
tmean=meanc(tmp);
tstd=stdc(tmp);
tmin=minc(tmp);
tmax=maxc(tmp);
tmed=medianx(tmp);
tq1=quantile(tmp,25);
tq3=quantile(tmp,75);
tp5=quantile(tmp,5);
tp95=quantile(tmp,95);
summary=tmean~tstd~tmin~tp5~tq1~tmed~tq3~tp95~tmax;
summary=summary';
clear tmp, tmean, tstd, tmin, tmax, tmed, tq1, tq3, tp5, tp95;

@print summary statistics for rate, rsq, ssres, gamma @
namedata; " "; " ";
"Descriptive statistics and quantiles"; " ";
let names=mean std min p5 q1 med q3 p95 max;
summary=names~summary;
let mask3[1,5]= 0 1 1 1 1;
let fmt3[5,3]=
"-*.*s " 8 8
"*.*lf " 8 3
"*.*lf " 8 3
"*.*lf " 8 3
"*.*lf " 8 3;
" "; "                  rate      rsq     ssres    gamma";
p=printfm(summary,mask3,fmt3); " ";
clear summary;

@ print highest and lowest rates and rsq's and corresponding id's @

"Cases with extreme values of rate and rsq, identified by id"; " ";
"Highest"; " ";
extrate=beta_1[.,2]~id;
extrsq=rsq_1~id;
extrate=rev(sortc(extrate,1));
extrsq=rev(sortc(extrsq,1));

let mask4[1,4]=1 1 1 1;
let fmt4[4,3]=
"*.*lf " 8 3
"*.*lf " 8 0
"*.*lf " 15 3
"*.*lf " 8 0 ;

"    rate         id                rsq          id";
 np < 20;
```

```
extremes=extrate~extrsq;
p=printfm(extremes,mask4,fmt4);    " ";
else;
hiextrms=extrate[1:10,.]~extrsq[1:10,.];
loextrms=extrate[np-10:np,.]~extrsq[np-10:np,.];
p=printfm(hiextrms,mask4,fmt4); " ";
p=printfm(loextrms,mask4,fmt4); " ";
endif;
clear extrate, extrsq, hiextrms, loextrms;
"Lowest"; "\f";

@create gauss data file with relevant variables @
if nl == 0;
tpmat=id~x~beta_1[.,2]~rsq_1~ssres_1~gam_i;
elseif nl > 0;
tpmat=id~x~omg~beta_1[.,2]~rsq_1~ssres_1~gam_i;
endif;

@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@
@ trimming                                                              @
@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@@

#include trimming
clear result, c_trm, d_rsq, gam_i, id, rsq_1, r_trm, ssreg_1, ssres_1;


ntrm:
#include psyc.prg;
output off;
if outfile == 1;
savextod;    @invokes interactive program to save output to
             a disk file.  The user supplies the name of the output
             file, the namo of the matrix in memory that is to be
             stored in the output file, and the name of a character
             variable, which names the columns (i.e., variables) in the
             matrix.  The name of the output file can be chosen arbitrarily.
             The name of the matrix in memory that will be stored is
             "gaussdat".  "varnames" is the character variable
             that supplies the variable names for the output file  @
endif;
end;
stoptp:;
end;
```

```
@ interactive version of timepath @
" "; " "; " "; " "; " "; " ";
"This is the Timepath program for fitting straight-line growth";
"curves to individual longitudinal data sequences.           ";
"This interactive question and answer session allows you to  ";
"specify the necessary information for running the program";
" ";
"First, you must describe the input file: ";
"       the matrix loaded will be np(# persons) x     ";
"       1(id) + nt(# waves) + nl(# correl.)          ";
"       You must specify the values np, nt, nl when prompted below. ";
"       The input file may be either an ASCII file (if <64k) or a ";
"       GAUSS data set.  In either case, the";
"       variables in the file are assumed to be in the following order:";
"       id  x1 x2 . . . xT w1 w2 . . . .                              ";
" ";
" ";
"If your data are not of this form, you must exit now and pre-process";
"the data before running this version of Timepath.";
" ";
"Are you ready to proceed?  (y, n)";
proceed=cons; " ";
if (proceed $== "n" or proceed $== "no");
goto stoptp;
endif;

"Are your data in an ASCII file or a GAUSS data set? (1=ASCII, 2=GAUSS)";
d=con(1,1); " ";

"Please give a name to your data that can be used in page";
"headings.  Use 40 characters or less."; " ";
"Title for data set:";;
namedata=cons; " "; " ";

"Enter the number of observations (i.e., individuals, cases):";
np=con(1,1);  " "; " ";

"Enter the number of waves of data for each individual;";
nt=con(1,1);       " "; " ";

"Enter the time points, e.g. 0 1 2 3, at which measurements were taken:";
tm=con(nt,1);      " ";     " ";

"Enter the number of background variables (w's) in the input file:";
nl=con(1,1);            " ";      " ";

if nl == 0;
c_of_c=0;
goto skip;
endif;

"Do you want to use these background variables to compute correlates";
"of change in this execution of Timepath?  (1=yes, 0=no)";
c_of_c=con(1,1); " "; " ";

skip:

"You can trim observations from the data automatically based";
```


BEST COPY AVAILABLE

```
"manually by specifying the id numbers of cases to be trimmed.";
"The trimmed cases are included in descriptive growth analyses";
"but are set aside before computing the psychometric quantities.";
"What type of trimming do you want to do?";
"(0=none, 1=automatic, 2=manual)";
trm=con(1,1); " ";

if trm == 1;

"Enter the percent of cases to be trimmed (a whole number >= 0):";
p_trm=con(1,1); " ";    " ";

elseif trm == 2;

"How many cases do you want to trim?";
n_trm=con(1,1); " ";

"Enter the id number of each case that you want to trim, followed by";
"a carriage return.";
c=1;
do while c <= n_trm;
tmpid=con(1,1);
if c == 1; selids=tmpid;
else;
selids=selids|tmpid;
endif;
c=c+1;
endo;
 " ";

endif;

"Do you want to create a GAUSS data set containing the scores,";
"estimated rates, and other information?   (1=yes, 0=no)";
outfile=con(1,1);  " ";    " ";

if outfile == 1;
"When the Timepath output is finished, you will automatically be";
"thrown into a menu-driven program that creates the GAUSS data set.";
"In order to complete the process, you will have to specify three ";
"things.  First, you must give a file name (including disk drive";
"specification) where the GAUSS data set will be stored.";
"The name is arbitrary.  Choose one that is meaningful to you.";
"Second, you must specify the GAUSS matrix in memory, which contains";
"the scores and statistical quantities that are to be saved in the";
"GAUSS data set.  Finally, you must specify a vector that contains";
"the names of the variables (or columns) in the data set.  The";
"GAUSS matrix and the vector are determined in the Timepath program,";
"and are as follows:";
" ";
"The data matrix to be stored is called 'tpmat'.";
"The variable (colunm) names are contained in the vector, 'varnames'.";
" ";
"You must remember 'tpmat' and 'varnames', as you will";
"be prompted for them when you create the GAUSS data set.";
"                         **********************";
" ";
endif;


"Please give the name of the file that contains the input data.";
"Use the form, d:filename.ext, where 'd' is the appropriate disk drive.";
" ";
"Input file: ";;
file=cons; " ";    " ";
```

```
"Give the name of the output file that will contain the hard copy";
"report produced by the Timepath program.  Be sure to give the drive";
"designation and the full file name.";
" ";
"output file: " ;;
hardcopy=cons;    " ";    " ";
```


ERIC

```
if trm == 0; p_trm=0; goto ntrm;
elseif trm == 2; goto st;
endif;

if p_trm <= 0;
   goto ntrm;
endif;
p_trm=p_trm./100;
n_trm=cint(np.*p_trm);
if n_trm == 0;
   " ***** percentage of trimming too small -- no cases trimmed ***** ";
   goto ntrm;
endif;
np_o=np;
np=np-n_trm;
c_trm=ssres_1~seqa(1,1,np_o);
c_trm=sortc(c_trm,1);
s_trm=c_trm[np+1:np_o,2];
r_trm=c_trm[1:np,2];

st:
if trm == 2;

n_trm=rows(selids);
np_o=np;
np=np-n_trm;
fullmat=data[.,1]~seqa(1,1,np_o);
si=1;
do while si <= n_trm;
  e=(fullmat[.,1] .== (selids[si,1].*ones(np_o,1)));
  if si == 1;
    c_trm=selif(fullmat,e);
    r_trm=delif(fullmat,e);
    e2=(r_trm[.,1] .== (selids[si+1,1].*ones(np_o-si,1)));
  else;
    c_trm=c_trm|selif(fullmat,e);
    r_trm=delif(r_trm,e2);
    if si < n_trm;
    e2=(r_trm[.,1] .== (selids[si+1,1].*ones(np_o-si,1)));
    endif;
  endif;
si=si+1;
endo;
s_trm=c_trm[.,2];
r_trm=r_trm[.,2];
clear fullmat, c_trm, e, e2;

p_trm=(n_trm./np_o);

endif;

@trimmed cases @
@
t_x=submat(x,s_trm,0);
t_beta_1=submat(beta_1,s_trm,0);
t_ssres=submat(ssres_1,s_trm,0);
t_ssreg=submat(ssreg_1,s_trm,0);
t_sstot=submat(sstot,s_trm,0);
```

```
t_rsq_1=submat(rsq_1,s_trm,0);
t_rsq_2=submat(rsq_2,s_trm,0);
t_drsq=submat(d_rsq,s_trm,0);
t_f_v=submat(f_v,s_trm,0);
t_f_s=submat(f_s,s_trm,0);
if c_of_c > 0;
   t_omg=submat(omg,s_trm,0);
endif;
@
@print cases deleted  @

namedata; " "; " ";
if trm == 1;
"Cases deleted due to high sums of squared residuals "; " ";
elseif trm == 2;
"Cases deleted manually"; " ";
endif;
if (nl >= 1 and c_of_c > 0);
idlbl$+wlbl$+stlbl2$+xlbl; " ";
else;
idlbl$+stlbl2$+xlbl;  " ";
endif;
p=printfm(submat(result,s_trm,0),mask2,fmt2);  " "; "\f";

@reset values of original variables so that they contain
 the cases retained for mle analysis                          @

x=submat(x,r_trm,0);
beta_1=submat(beta_1,r_trm,0);
ssres_1=submat(ssres_1,r_trm,0);
@
ssreg_1=submat(ssreg_1,r_trm,0);
sstot=submat(sstot,r_trm,0);
rsq_1=submat(rsq_1,r_trm,0);
rsq_2=submat(rsq_2,r_trm,0);
d_rsq=submat(d_rsq,r_trm,0);
f_v=submat(f_v,r_trm,0);
f_s=submat(f_s,r_trm,0);
@
if (nl >= 1 and c_of_c > 0);
   omg=submat(omg,r_trm,0);
endif;
```

```
@ Blomquist estimation                                                    @

@ sum of squares for time                                                 @

ss_tm=sumc(tm.^2)-nt*((meanc(tm)).^2);

@ parameter covariance matrix                                             @

beta_d=beta_1- ones(np,1)*meanc(beta_1)';
cv_beta=(beta_d'beta_d)./np;

@ blomquist estimates of covariance matrix of initial status and rate
  of change                                                               @

err_var=sumc(ssres_1)./(np*(nt-2));
ins_var=cv_beta[1,1]-((err_var.*sumc(tm.^2))./(ss_tm.*nt));
trt_var=cv_beta[2,2]-(err_var./ss_tm);
ist_cov=cv_beta[2,1]+(err_var.*meanc(tm)./ss_tm);

@ estimation of t-zero, variance of ksi(t-zero) and the scaling
  factor                                                                  @

t_zero=-1.*(ist_cov./trt_var);
v_ksi_t0=((ins_var.*trt_var)-(ist_cov.^2))./trt_var;
scl_fct=sqrt(v_ksi_t0./trt_var);
sem_rate=sqrt(cv_beta[2,2]-trt_var);

@ estimation of variance(ksi(t)) and the covariance(ksi(t),t_rate         @

t_tzr=tm-ones(nt,1).*t_zero;
c_ksitr=(ones(nt,1).*trt_var).*(t_tzr);
v_ksi_t=(ones(nt,1).*v_ksi_t0)+(ones(nt,1).*trt_var).*(t_tzr.^2);

@ estimation of regression coefficient and corr of t_rate on ksi(t)      @

b_trksi=c_ksitr./v_ksi_t;
r_trksi=c_ksitr./sqrt(ones(nt,1).*trt_var.*v_ksi_t);

@ observed variance and between-waves correlation--trimmed data  @

x_bar=meanc(x);
x_dev=x-ones(np,1)*x_bar';
cov_x=(x_dev'*x_dev)./np;
sd_x=sqrt(diag(cov_x));
corr_x=cov_x./(sd_x*sd_x');

@print observed correlation matrix--trimmed  @

namedata; " "; " ";
"Between-wave correlation matrix--trimmed data ";
format /m1 /rd 8,3;
tm1=tm+i_tm*ones(nt,1);
"            " tm1';
"              " tm1~corr_x;
"\f";

@observed variance and reliability of fitted values   @
fits=beta_1*t_1';
```

```
f_bar=meanc(fits);
f_dev=fits-ones(np,1)*f_bar';
cov_fit=(f_dev'*f_dev)./np;
var_fit=diag(cov_fit);
rel1_fit=v_ksi_t./var_fit;

@ reliability estimation of trate_hat and x(i)                          @

rel_tr=trt_var./cv_beta[2,2];
rel_xt=v_ksi_t./(sd_x^2);
rel_ksi=corr_x./sqrt(rel_xt*rel_xt');
rel_ksi=diagrv(rel_ksi,ones(nt,1));
t_zero_s=t_zero+i_tm;
@ correlates of change statistics                                       @

if (nl == 0 or c_of_c == 0);
  goto xxx;
endif;

@ additional covariance estimates involving omega                       @

omg_dev=omg-ones(np,1)*meanc(omg)';
pv=beta_d~omg_dev;
c_pv=(pv'*pv)./np;
v_omg=diag(c_pv);
v_omg=v_omg[3,nl+2,.];
r=seqa(3,1,nl)';
c_th_w=submat(c_pv,r,2);
c_ks0_w=submat(c_pv,r,1);
c_kst0_w=c_ks0_w-((c_th_w.*ist_cov)./trt_var);

@ correlations and regression estimates                                 @

r_th_w=c_th_w./sqrt(trt_var.*v_omg);
b_th_w=c_th_w./v_omg;
r_kst0_w=c_kst0_w./sqrt(v_ksi_t0.*v_omg);
b_kst0_w=c_kst0_w./v_omg;
t_u=t_zero.*ones(nl,1)+((v_ksi_t0.*c_th_w)./(trt_var.*c_kst0_w));
t_l=t_zero.*ones(nl,1)-(c_kst0_w./c_th_w);
t_u_s=t_u+i_tm;
t_l_s=t_l+i_tm;

@ compute wave by wave correlations                                     @

c_x_w=(omg_dev'*x_dev)./np;
r_x_w=c_x_w./sqrt((ones(nl,1)*(sd_x^2)').*(v_omg*ones(1,nt)));
b_x_w=c_x_w./(v_omg*ones(1,nt));
b_ksi_w=(b_kst0_w*ones(1,nt))+((b_th_w*ones(1,nt)).*
                              (ones(nl,1)*t_tzr'));
r_ksi_w=r_x_w./sqrt(ones(nl,1)*rel_xt');
xxx:
@ output code follows here                                              @
format /m1 /rd 8,3;
"                   Time Path on " namedata;
" ";
"           Foulkes-Davis Tracking Index";" ";
"           Gamma:                          " gam_fd;
"           Standard Error (Gamma):    " se_gam;" ";" ";
"           Blomqvist Maximum Likelihood Estimates";" ";
if trm == 0;
"           Percent of cases trimmed:" p_trm.*100; " ";
elseif trm == 1;
" Percent of cases automatically trimmed:" p_trm.*100; " ";
elseif trm == 2;
" Percent of cases manually trimmed:" p_trm.*100; " ";
```

```
"           t(zero):                  " t_zero_s;
"           Scaling Factor:           " scl_fct;
"           True Rate Variance:       " trt_var;
"           Variance Ksi[t(zero)]:    " v_ksi_t0;
"           Reliability of Rate:      " rel_tr;
"           S. E. Measurement of Rate:" sem_rate;" ";
tm=tm+i_tm*ones(nt,1);
format /m1 /rd;
"        Disattenuated Correlation Matrix";
"           " tm';
"            " tm~rel_ksi;" ";" ";
"                    " tm';" ";
"   Var(Ksi)          " v_ksi_t';
"   Var(X)            " diag(cov_x)';
"   Var(X_hat)        " var_fit';
"   Reliability(X)  " rel_xt';
"   Reliab(X_hat)   " rel_fit';
"   Beta(Rate,Ksi)  " b_trksi';
"   Corr(Rate,Ksi)  " r_trksi';
" ";" ";" ";
"\f";

if (nl == 0 or c_of_c == 0);
goto last;
endif;

"           Systematic Individual Differences";" ";
crlt=seqa(1,1,nl);
"   Correlate ->    " crlt';" ";
"   Corr(Rate,W)    " r_th_w';
"   Beta(Rate,W)    " b_th_w';
"   Corr(Ksi_t0,W)  " r_kst0_w';
"   Beta(Ksi_t0,W)  " b_kst0_w';
"   t(Upper)         " t_u_s';
"   t(Lower)         ' t_l_s';" ";" ";
"          Correlation Matrix of X(t) <columns> and W <rows>";" ";
"   Waves ->         " tm';
"       Correlate";
"  "  crlt~r_x_w;" ";" ";
"          Correlation Matrix of Ksi(t) <columns> and W <rows>";" ";
"   Waves ->         " tm';
"       Correlate";
"  " crlt~r_ksi_w;" ";" ";
"          Regression Coefficients of X(t) <columns> on W <rows>";" ";
"   Waves ->         " tm';
"       Correlate";
"  " crlt~b_x_w;" ";" ";
"          Regression Coefficients of Ksi(t) <columns> on W <rows>";" ";
"   Waves ->         " tm';
"       Correlate";
"  " crlt~b_ksi_w;" ";
last:;
```